Tampereen yliopisto

Sami Rustholkarhu

IHMINEN VASTAAN KONE

Milloin tekoäly erehtyy?

Johtamisen ja talouden tiedekunta
Pro gradu -tutkielma
Ohjaaja: Kari Lohivesi
Huhtikuu 2019

TIIVISTELMÄ

Sami Rustholkarhu: Ihminen vastaan kone – Milloin tekoäly erehtyy?
Pro gradu -tutkielma
Tampereen yliopisto
Yrityksen johtaminen
Huhtikuu 2019

Tässä tutkimuksessa kuvataan ja analysoidaan tekoälyä päätöksenteon kontekstissa. Tutkimus kuvaa tekoälyä päätöksentekijänä, esittelee ja analysoi tilanteita, joissa tekoäly ei toiminut odotetulla tavalla, sekä suhteuttaa tekoälyn toimintaa inhimilliseen päätöksentekoon.

Tutkimuksen kirjallisuus rakentuu inhimillistä päätöksentekoa kuvaavien tutkimusten varaan. Kirjallisuus lähestyy päätöksentekoa sekä ihmisen rajoitteiden että heuristisen päätöksenteon tehokkuuden kautta ja esittelee myös organisaatioviitekehykselle ominaisia päätöksenteon piirteitä. Tutkimuksen teoreettinen viitekehys nivoo kirjallisuudesta nostetut kokonaisuudet yhteen inhimillisen päätöksenteon prosessimaisella mallintamisella sekä käsittelemällä päätöksentekoa ympäristöstä muodostettujen säännönmukaisuuksien varaan rakentuvan intuition kautta.

Tämän laadullisen tutkimuksen empiirinen aineisto muodostettiin triangulatorisesti useita eri lähteitä ja menetelmiä hyödyntäen. Tutkimuksen tekoälyn päätöksentekoa kuvaava primaariaineisto muodostettiin tekoälyteknologian kehittäjiltä ja alan tutkijoilta kerätyistä teemahaastatteluista. Tekoälyn erehtyväisyyttä lähestyttiin uutisartikkeleista, yritysten tutkimusraporteista sekä blogijulkaisuista muodostetun sekundaariaineiston kautta esittelemällä julkisuuteen nousseita tapauksia tekoälyn tekemistä virheistä.

Tutkimus havaitsi intuition käsitteen olevan vahvasti tekoälyä ja inhimillistä päätöksentekoa yhdistävä kokonaisuus. Vaikka intuitiivinen, ympäristöstä havaittavia säännönmukaisuuksia korostava, päätösprosessi on ominaista sekä ihmiselle, että tekoälylle eivät tekoälyn päätöksentekoa kuvaavat piirteet jaa muuten yhtäläisyyksiä inhimillisen päätöksenteon kanssa. Tutkimusaineistosta löydettiin kolme tekoälyn päätöksentekoa kuvaavaa kokonaisuutta: kohdennettu soveltaminen, rajoitettu deskriptiivisyys ja riippuvuus kolmansista osapuolista. Tekoäly toimii ympäristön, suoritettavan tehtävän ja käytettävän opetusdatan suhteen aina äärimmäisen kapealla alueella eikä sen päätöksenteko useissa tapauksissa ole rationalisoitavissa. Tekoälyn antamaa tulkintaa voidaan verrata haluttuun lopputulokseen, mutta kyseenalaistettavissa olevaa päättelyketjua ei voida tekoälyn tapauksessa usein esittää. Tekoälyn toiminta on myös äärimmäisen riippuvaista sitä ympäröivistä ulkopuolisista tekijöistä. Sekä opetusdatan valinta että siitä tekoälyn tekemät tulkinnat ovat osin ulkopuolisen toimijan, usein ihmisen, ohjaamia.

Tekoälyn erehtyväisyyden näkökulmasta tutkimus korostaa tekoälyn päätöksentekoa kuvaavien piirteiden suhdetta sille annettuun tehtävään ja ympäristöön, jossa tehtävää suoritetaan. Ihmisen päätöksentekoon verrattavissa olevaa systemaattista päätöksenteon vinoutumista ei tekoälyn tapauksessa esiinny, vaan mahdolliset päätöksenteon virheet aiheutuvat teknologian, ympäristön ja suoritettavan tehtävän epäonnistuneesta yhteensovittamisesta.

Tutkimustulosten voidaan katsoa korostavan teknologiavuorovaikutuksen merkitystä. Useista muista teknologioista poiketen vuorovaikutus ei ainoastaan vaikuta teknologiasta saatavaan hyötyyn, vaan muuttaa merkittävällä tavalla teknologian toimintaa. Tämä tarkoittaa, että tekoälyavusteista päätöksentekoa ymmärtääksemme emme voi tarkastella tekoälyä erillisenä toimijana, vaan meidän tulee suunnata huomiomme teknologiavuorovaikutukseen sekä kontekstiin, jossa tämä vuorovaikutus tapahtuu.

Avainsanat: Päätöksenteko, Vinoumat, Heuristiikat, Tekoäly, Koneoppiminen

SISÄLLYS

Kuviot

Taulukot

1 JOHDANTO

1.1 Tutkimusaiheena ihmisen ja koneen erehtyväisyys

Inhimillinen erehtyväisyys on tunnustettu päätöksentekokirjallisuudessa 1950-luvulta lähtien, kun Simon (1957) julkaisi rajoitetun rationaalisuuden käsitteen. Tämän jälkeen päätöksentekoa käsittelevä tutkimus on keskittynyt päätöksenteon vinoumien ja ihmisen erehtyväisyyden selittämiseen (Arkes ja Blumer 1985; Kahneman ja Tversky 1974; 1979; 1981; 1986; Simon 1955; 1956, 1957; Staw 1976; 1981; Staw ja Fox 1977; Staw ja Ross 1978; Whyte 1986). Aihetta on lähestytty myös päätöksentekoympäristöjen (Simon 1956, Gigerenzer 2008) ja heuristiikkojen toimivuuden (Goldstein ja Gigerenzer 1996; 2002; Jacoby ja Dallas 1981; Johnson ja Goldstein 2004) näkökulmista. Rationaalisesti rajoittuneelle päätöksenteolle on tyypillistä yksinkertaistaminen sekä valitun lopputuloksen että päätöksentekoympäristön suhteen. Olemme taipuvaisia lopettamaan uusien vaihtoehtojen etsimisen sekä vaihtoehtojen välisen vertailun ensimmäisen tavoitetason tyydyttävän vaihtoehdon tultua vastaan. Kognitiiviset rajoitteemme taas estävät huomioimasta ympäristön kaikkia lopputulokseen vaikuttavia muuttujia tai muuttujien välisiä painoarvoja. (Simon, 1955; 1956; 1957) Nämä tavoitteita ja ympäristöä koskevat yksinkertaistukset tekevät meistä toisinaan haparoivia, yliluottavaisia ja virhealttiita (Kahneman 1974; 1979; 1981; 1986). Toisaalta ne ovat tehneet päätöksenteostamme myös nopeaa ja tehokasta (Gigerenzer 2008; Gigerenzer ja Brighton 2009). Ihmislajin tehokasta levittäytymistä voidaan pitää eräänlaisena todisteena siitä, että olemme onnistuneet optimoimaan päätöksentekokoneistomme siten, että se pystyy saatavilla olevalla laskentateholla tuottamaan, valtavassa määrässä erilaisia päätöksentekoympäristöjä, tarpeen tyydyttämiseen riittävän tuloksen nopeasti ja riittävällä tarkkuudella.

Tilanteissa, joissa meidän tulisi tietyssä määrätyssä ympäristössä saavuttaa tulos, joka ei olisi pelkästään riittävän hyvä vaan optimaalinen, olemme pitkään turvautuneet koneiden apuun. Kone on, aikaisimpia mekaanisia laskukoneita lukuun ottamatta, aina ollut ihmistä tehokkaampi toistuvissa tilastollista päättelyä edellyttävissä tehtävissä (Nilsson 2010). Olemmekin pyrkineet ulkoistamaan näitä kognitiivisesti raskaita, aikaa vieviä ja itseään toistavia tehtäviä koneellisten toimijoiden suoritettavaksi mahdollisimman tehokkaasti.

Tämä tehtävien ulkoistus on edellyttänyt myös muutoksia toimintatapoihimme. Koska teknologia ei ole juurikaan sopeutunut muihin ympäristöihin, kuin mihin se on alun perin luotu, on meidän täytynyt muuttaa itseämme ja toimintaamme konetta paremmin palveleviksi. Uuden teknologian toimintaamme muokkaava vaikutus ei ole vain tälle ajalle tai digitalisaatiolle ominaista. Harari (2015) kuvailee, kuinka maanviljelyyn siirtyneet metsästäjäkeräilijät alkoivat noin 8500 eaa. järjestäytyä suuremmiksi yhteisöiksi ja eriyttää tehtäviä pystyäkseen tehokkaammin hyödyntämään juuri domestikoitujen viljakasvien kasvattamista edesauttavia teknologioita. Tällainen ulkoisesta toimijasta johtuva muutos käyttäytymisessä ja työtavoissa on toistunut historiamme aikana aina kun uusi tarpeeksi merkittävä teknologia on esitelty (Harari, 2015). Olemme pyrkineet koneenkaltaistamaan ajatteluamme ja toimintaamme, jotta saisimme keksimästämme koneesta suurimman mahdollisen potentiaalin irti.

Noin 10 vuotta ennen inhimillisen erehtyväisyyden löytämistä neuropsykologi Warren McCulloch ja matemaatikko Walter Pitts (1943) olivat esittäneet matemaattisen mallin ihmisen aivoissa olevan neuronin toiminnasta ja osoittaneet, että tällaisista keinotekoisista neuroneista muodostettu verkko kykenisi suoriutumaan kaikista mahdollisista laskutoimituksista. Tätä voidaan pitää eräänä merkittävänä esiaskeleena myöhemmin tekoälytutkimukseksi nimetyn tieteenalan synnyssä. Omaksi tutkimusalueekseen eriytymisen jälkeen tekoäly on elänyt rationaalisesti rajoittuneen ihmisen rinnalla sekä popkulttuurissa että tieteellisessä tutkimuksessa kokien sekä nousuja että laskuja. Juuri nyt elämme tekoälyhypen keskiössä. Louridas ja Ebert (2016) kuvailevat teollisen vallankumouksen veroista arvon tuottamisen ja kuluttamisen mullistusta, jossa tekoäly tulee näyttelemään merkittävää roolia. Emergenteille teknologioille asetettuja odotuksia mittaavalla Gartnerin hype-käyrällä tekoälyn ja koneoppimisen sovellukset ovat sijoittuneet käyrän huipulle vuodesta 2016 lähtien (Gartner 2016; 2017; 2018).

Käsitteenä tekoäly on hatarasti määritelty kattotermi teknologioille, joiden sovellusalueet ovat koneille äärimmäisen haastavia. Laajasti määriteltynä tekoälyn voidaan katsoa pitävän sisällään kaikki aktiviteetit, joiden pyrkimyksenä on koneiden älykkääksi tekeminen (Nilson 2010, 13). Tekoälyn sovellusalueille on yhteistä, että ne ovat kokonaisuuksia, joilla ihminen on pitkään suoriutunut konetta tehokkaammin. Esimerkiksi erilaisten objektien tunnistaminen ja erottaminen toisistaan on ollut haaste, jossa ihminen on aiemmin suoriutunut ylivertaisesti koneeseen verrattuna (Nilson 2010).

Tekoäly jakaa McCullochin ja Pittsin (1943) ansiosta tietyn neurotieteellisen ja biologisen pohjan ihmisen toiminnan kanssa. Tämän yhdistävän tekijän ja toimintaamme imitoivien sovellusalueiden vuoksi tekoäly näyttäytyy ennennäkemättömänä mahdollisuutena inhimillistää koneemme. Tämä antaa orastavan lupauksen siitä, että – toisin kuin Hararin (2015) kuvailemat 8500 eaa. maanviljelyyn siirtyneet metsästäjäkeräilijät ja heitä seuranneet sukupolvet – meidän ei enää tarvitsisi taipua teknologian rajoittaviin toimintamalleihin, vaan teknologia sopeutuisi meille ominaiseen käyttäytymiseen. Tämä toistaiseksi suurilta osin realisoitumaton lupaus ei kuitenkaan anna meille tarpeeksi työkaluja analysoida ihmisen ja koneen tulevaisuuden yhteistyöpotentiaalia. Emme osaa ennustaa, riittääkö neurobiologinen pohja, ja ihmistoimintaa muistuttavat sovellusalueet sopeuttamaan teknologian meidän toimintaamme. Emmekä osaa sanoa, aiheuttavatko teknologian inhimillistämisyritykset meille ominaisen erehtyväisyyden ennennäkemättömän skaalautumisen.

1.2 Tutkimuskysymykset

Tämä tutkimus suhteuttaa tekoälyn päätöksentekoa inhimilliseen päätöksentekoon. Päätöksenteon viitekehyksellä tutkimus rakentaa siltaa organisaatio- ja tekoälytutkimuksen välille tarjoamalla käytännöllisiä ja teoreettisia näkökulmia sekä teknologian kehittämiseen että sen soveltamiseen. Tutkimuksen tavoitteena on ihmisen toimintaan suhteuttaen ymmärtää tekoälyn päätöksentekoa ja erehtyväisyyttä. Onnistuakseen tavoitteessaan tutkimuksen tulee:

1) Kuvata inhimillisen päätöksenteon erityispiirteitä
2) Kuvata tekoälyn päätöksenteon erityispiirteitä
3) Kuvata ja analysoida tekoälyn tekemiä virheitä

Tutkimuksen tavoitteisiin vastataan kirjallisuuden, vapaasti saatavien verkkoaineistojen, kuten uutisten, blogien ja keskustelufoorumijulkaisuiden sekä asiantuntijahaastatteluiden avulla. Kuvaus inhimillisestä päätöksenteosta muodostetaan kirjallisuutta hyödyntäen. Koneellisen päätöksenteon sekä tekoälyn tekemien virheiden kuvaus rakennetaan tutkijan keräämän primaari ja sekundaariaineiston avulla.

1.3 Tutkimuksen rajaus

Tutkimus vertailee inhimillistä päätöksentekoa tekoälyjärjestelmien päätöksentekoon. Erehtyväisyys on valittu vertailun lähtökohdaksi, sillä se on hyvin edustettu teema päätöksentekoa kuvaavassa kirjallisuudessa tarjoten laajan, tiedeyhteisön validoiman kirjallisuuden yli 50 vuoden ajalta. Inhimillistä päätöksentekoa käsittelevä kirjallisuus on tutkimuksen tavoitteen ehdoilla rajattu 1950-luvun jälkeiseen klassisen rationaalisuuden käsitteen kyseenalaistavaan ja ihmisen tiedonkäsittelykykyjen rajallisuuden tunnustavaan kirjallisuuteen.

Tutkimus tiedostaa neurotieteiden ja biologian kontribuution inhimillisen päätöksenteon kuvaamisessa muttei katso niiden palvelevan tutkimuksen tavoitteen määräämää rajausta. Tutkimuksen ydintavoitteen näkökulmasta nähdään tärkeämmäksi kuvata yksilön päätöksentekoa suhteessa ympäristöön, kuin selittää päätöksentekoa yksilön ominaisuuksilla. Neurotieteisiin verrattuna päätöksentekokirjallisuuden katsotaan myös paremmin tarjoavan uusia näkökulmia tekoälytutkimukseen siirtämällä vertailun painopisteen hermostotason biologisesta tarkastelusta ja matemaattisista malleista lopputuloksen tarkasteluun. Uskon vankan, hyvin argumentoidun kirjallisuuspohjan myös liittävän tekoäly tiukemmin kauppatieteellisen tutkimusdiskurssin ytimeen tarjoten teknologian soveltajille helpommin käsitteellistettävän toimintaympäristön.

Inhimillisen päätöksenteon tapauksessa neurotieteellisen näkökulman rajaaminen tutkimuksen ulkopuolelle on mahdollista, sillä tutkimuskysymyksen asettelun kannalta relevantimpaa ja vahvasti argumentoitua kirjallisuutta on löydettävissä. Tekoälyä ei kuitenkaan ole kirjallisuudessa aiemmin käsitelty päätöksenteon näkökulmasta ihmisen päätöksentekoa vastaavalla laajuudella. Tämän vuoksi tutkimus ei voi täysin irrottautua yksittäisten mallien mikrotason kuvaamisesta tekoälyn tapauksessa, vaikka tekeekin sen inhimillisen päätöksenteon tapauksessa. Tutkimuksen tavoitteen kannalta ei kuitenkaan nähdä tarkoituksenmukaiseksi listata yksityiskohtaisesti tilastollisia malleja tai esittää määrittelyä siitä, mitkä menetelmät voidaan katsoa kuuluvan tekoälyn käsitteen alle ja mitkä ei. Tilastollisten ja matemaattisten mallien esittelyllä pyritään rakentamaan ymmärrystä tekoälystä päätöksenteon näkökulmasta. Malleja käytetään eräänä työkaluna havainnollistamaan sitä, millaista tekoälyn suorittama päättely on ja niiden käsittelyn laajuus on rajattu palvelemaan tätä tavoitetta. Tutkimusraportissa malleista pyritään nostamaan esiin vain niiden päätöksentekoa ilmentävät piirteet. Mallien kuvaukset esitetään tutkimusraportissa liitteinä.

Tekoäly käsitetään tutkimuksessa lukuisia teknologioita sisältävänä kattoterminä. Tutkimus ymmärtää siten läheisyytensä tekoälyn alla kulkevien lukuisten menetelmällisten käsitteiden, kuten siirto- ja vahvistusoppimisen tai rekursiivisten ja konvolutionaalisten neuroverkkojen, kanssa. Tekoälyn suorittaman päätöksenteon kuvaamisen näkökulmasta näiden ei kuitenkaan katsota muodostavan kriittistä kokonaisuutta, joten niiden käsittely on rajattu tutkimuksen ulkopuolelle tutkimuksen laajuuden hillitsemiseksi. Tutkimuskysymyksen näkökulmasta ei myöskään nähdä tarpeelliseksi keskustella liitteissä esiteltyjen tilastollisten mallien suhteesta menetelmällisiin käsitteisiin.

Kuten tutkimuksen kirjallisuus, myös empiria on rajattu palvelemaan päätöksenteon kuvausta. Empirian ydintehtävä on kuvata tekoälyn päätöksentekoa ja erehtyväisyyttä inhimilliseen päätöksentekoon verrattavissa olevalla tavalla. Tämä tehdään esittelemällä julkisuuteen nousseita tapauksia tekoälyjärjestelmien tekemistä virheistä ja syventämällä virheiden analysointia muun tutkimusaineiston avulla. Tutkimusaineistoon kuuluvissa haastatteluissa kaikki haastatteluihin osallistuneet ovat joko tekoälytutkijoita, teknologian kehittäjiä tai sen jollain sovellusalueella työskenteleviä henkilöitä, joilla katsotaan olevan asiantuntijuuden edellyttämät tiedot tekoälystä tai sen hyödyntämisestä. Haastatteluaineisto nostaa esille satunnaisia menetelmiin liittyviä teknisiä käsitteitä, jotka avataan lyhyesti sillä laajuudella kuin päätöksenteon näkökulmasta on välttämätöntä. Näiden lyhyiden kuvausten ei ole kuitenkaan tarkoitus olla yksityiskohtaisia läpikäyntejä menetelmistä tai niiden soveltamisesta.

1.4 Keskeiset käsitteet ja lyhenteet

Tässä kappaleessa määritellään lyhyesti tämän tutkimuksen kannalta keskeiset käsitteet. Käsitteet on määritelty tämän tutkimuksen näkökulmasta ja ne avataan tarkemmin joko tutkimuksen kirjallisuudessa tai tekoälyä käsittelevässä luvussa.

Ekologinen rationaalisuus – Inhimillistä päätöksentekoa kuvaava käsite, joka suhteuttaa yksilön tavoitetason yksilön toimintaympäristöön. Yksilö pyrkii toimimaan rationaalisesti suhteessa näihin. Esimerkiksi päätösstrategia valitaan ekologisen rationaalisuuden ohjaamana.

Heuristiikka – Psykologiassa käytetty termi, jolla tarkoitetaan karkeaa menetelmää, jota käytetään päätöksentekoon, ongelman ratkaisuun tai arvion muodostamiseen ilman

kaikkien tarjolla olevien vaihtoehtojen vertailua tai kaiken saatavilla olevan tiedon hyödyntämistä.

Koneoppiminen – Keino löytää datasta automaattisesti säännönmukaisuuksia

Luonnollisen kielen käsittely – Tekoälyn osa-alue, joka tarkastelee, kuinka koneita voidaan käyttää ymmärtämään luonnollista, ihmisten käyttämää, sekä puhuttua että kirjoitettua kieltä. Luonnollisen kielen käsitteellä erotetaan ihmisen puhumat kielet koneen ymmärtämistä kielistä

Päätösstrategia – Tietoisesti tai tiedostamatta valittu mekanismi, jonka mukaan päätös tietyssä ympäristössä tehdään

Rajoitettu rationaalisuus – Inhimillistä päätöksenteko kuvaava käsite, jonka mukaan ihminen ei tiedonkäsittelykykyjensä rajallisuuden vuoksi päätöksentekotilanteessa vertaile kaikkia vaihtoehtoja ja optimoi valintaansa, vaan valitsee ympäristöönsä ja tavoitetasoonsa suhteessa ensimmäisen riittävän hyvän vaihtoehdon

Tekoäly – Kattotermi lukuisille koneen älykkääksi tekemiseen käytetyille teknologioille.

Virhe – Toiminta, joka eroaa annetun tavoitteen mukaisesta, odotetusta toiminnasta

1.5 Tutkimusraportin rakenne

Tutkimusraportti on jaettu kuuteen lukuun. Johdantoluku esittelee tutkimuksen aihepiirin, listaa lyhyesti tutkimuksen kannalta keskeisimmät käsitteet sekä esittelee tutkimusongelman asettelun ja tutkimusta rajaavat valinnat. Toinen luku koostuu tutkimuksen kirjallisuudesta. Luku käy läpi Simonin (1957) rajoitetun rationaalisuuden käsitteen ympärille rakentunutta yksilön päätöksentekoa tarkastelevaa kirjallisuutta sekä esittelee päätöksentekoa organisaation viitekehyksessä. Tutkimuksen ensimmäiseen alatavoitteeseen kuvata inhimillisen päätöksenteon erityispiirteitä, vastataan toisen luvun lopussa esitetyillä inhimillistä päätöksentekoa kuvaavilla malleilla, joiden tehtävänä on myös nivoa tutkimuksessa käytetty kirjallisuus tutkimuksen tavoitteita palvelevaan muotoon.

Tutkimusraportin kolmannessa luvussa perustellaan tutkimukseen liittyviä metodisia valintoja. Luku ottaa kantaa tutkimusstrategiaan, aineiston keräämistä ja analysointia koskeviin seikkoihin sekä esittelee tutkimusprosessin etenemistä. Lisäksi luvussa

eritellään tutkimuksen luotettavuuteen vaikuttavia tekijöitä ja arvioidaan tutkimuksessa tehtyjen valintojen vaikutusta tutkimuksen kokonaisluotettavuuteen.

Luvussa neljä käsitellään tekoälyn päätöksentekoa. Luku esittelee osan tutkimuksen primaari- ja sekundaariaineistosta, määrittelee tekoälyn ja koneoppimisen käsitteet sekä kuvaa teknologian toimintaa. Luvun lopussa oleva tekoälyn päätöksenteon synteesi vastaa tutkimuksen toiseen alatavoitteeseen: kuvata tekoälyjärjestelmien päätöksenteon erityispiirteitä.

Viidennessä luvussa esitellään loput tutkimuksen aineistosta. Luku esittelee tekoälyn tekemät virheet tapaus kerrallaan ja analysoi niitä luvussa neljä esitetyn aineiston perusteella. Luku täyttää tutkimuksen tavoitteen kannalta viimeisen kriittisen alatavoitteen kuvaamalla ja analysoimalla tekoälyn tekemiä virheitä.

Tutkimusraportin viimeinen luku tarkastelee tekoälyn päätöksentekoa suhteessa ihmiseen sekä liittää tutkimuksen reaalimaailmaan pohtimalla automatisoidun ja inhimillisen päätöksenteon yhteistoimintaa. Lisäksi luvussa avataan näkemystä tutkimuksen käytännöllisen ja teoreettisen kontribuution merkityksistä organisaatio- ja tekoälytutkimukselle sekä avataan suuntia mahdolliselle jatkotutkimukselle.

2 PÄÄTÖKSENTEKO

Päätöksentekoa käsittelevä teoria on 50-luvulle asti noudattanut klassisen ja uusklassisen taloustieteen oletusta yksilön rationaalisesta hyödyn maksimoinnista ja kustannusten minimoinnista. Klassisen rationaalisuuden käsitteen mukaan päätöksentekijä vertailee eri vaihtoetoja niiden tuottaman loppuhyödyn perusteella ja valitsee vaihtoehdon, joka optimoi kustannus- ja hyötyfunktiot. Simonin (1955, 99) mukaan tällainen käsitys kuvaa heikosti reaalimaailman päätöksentekotilanteita, sekä asettaa päätöksiä tekevälle organismille vaatimuksia, joita ei ole mahdollista täyttää. Simonin (1957) esittelemä rajoitetun rationaalisuuden käsite on ollut pohjana myöhemmälle päätöksentekokirjallisuudelle

Tässä luvussa esitellään tutkimuksessa käytettävä kirjallisuus. Luvun tehtävänä on vastata tutkimuksen ensimmäiseen alatavoitteeseen, kuvata inhimillisen päätöksenteon ominaispiirteitä. Luvun rakenne etenee kronologisessa järjestyksessä alkaen 1950-luvun jälkeen julkaistusta taloustieteellisen klassisen rationaalisuuden käsitteen kyseenalaistavasta kirjallisuudesta ja siirtyy päätöksenteon vinoumien kautta intuitioon, asiantuntijuuteen ja heuristiseen päättelyyn. Luvun viimeinen kappale 2.9 tiivistää tutkimuksessa käytetyn kirjallisuuden esittämällä kaksi inhimillistä päätöksentekoa kuvaavaa mallia.

2.1 Rajoitettu rationaalisuus

Herbert Simon esitteli rajoitetun rationaalisuuden käsitteen vuonna 1957 julkaistussa artikkelissaan (Simon, 1957). Simonin (1955, 104) mukaan taloustieteisiin pohjautuva klassinen rationaalisuus edellytti, että päätöksentekijä (1) tietää jokaisen valittavissa olevan vaihtoehdon seuraukset, (2) tuntee tarkoin tavoiteltavan lopputuloksen ja valittavissa olevien vaihtoehtojen suhteen, (3) kykenee erottelemaan mahdolliset lopputulokset yksilöllisesti siten, ettei odottamattomille seurauksille jää tilaa, (4) pystyy asettamaan lopputulokset preferenssijärjestykseen siten, että niiden välinen vertailu on kaikissa tilanteissa mahdollista sekä (5) kykenee todennäköisyyksiin perustuvaan laskennalliseen vaihtoehtojen arviointiin. Simonin (1955, 104) mukaan ei ole empiirisiä todisteita, että ihminen täyttäisi klassisen rationaalisuuden käsitteen päättävälle organismille asettamat vaatimukset. Klassiseen rationaalisuuteen pohjautuvaa näkemystä ei siten voida pitää reaalimaailman päätöksentekoa kuvaavana (Simon 1955).

Simon (1955, 101) korostaa, että varsinkin laskennallisten ja ennustamista edellyttävien rajoitteiden vuoksi inhimillinen päätöksenteko reaalimaailman tilanteissa on parhaimmillaankin hyvin yksinkertaistettua klassiseen rationaalisuuden käsitteeseen verrattuna. Yksilöt harvoin tietävät valintojensa tarkkoja lopputuloksia tai tiettyjen lopputulosten todennäköisyyksiä. Lopputulosten preferenssijärjestys on usein osittainen ja ympäristön mukaan muuttuva. (Simon 1955) Simon (1955, 108) kyseenalaistaa myös klassisen rationaalisuuden oletuksen vaihtoehtojen vertailusta. Hän korostaa, että reaalimaailmassa vaihtoehdot näyttäytyvät päätöksentekijälle harvoin samanaikaisesti. Sen sijaan ne tulevat esiin satunnaisessa järjestyksessä. Päätöksentekijä ei siis valitse kaikista mahdollisista vaihtoehdoista parasta vaan ensimmäisen, joka vastaa päätöksentekijän sen hetkiseen tavoitetasoon. Tavoitetaso on päätöksentekijälle subjektiivinen ja vaihtelee ympäristön mukaan. Mikäli tavoitetasoon yltävien vaihtoehtojen löytäminen on helppoa, taso yleensä nousee, jos se on vaikeaa, taso laskee (Simon 1955, 111)

Rajoitettu rationaalisuus korostaa ympäristön merkitystä päätöstilanteessa. Ympäristöllä Simon (1956) viittaa päätöksentekijän ulkopuolisiin seikkoihin, jotka toimija näkee tilanteessa relevanteiksi. Tavoitetaso ja toiminnan ajurit ovat siis merkittäviä tekijöitä päätöksentekijälle relevantin ympäristön muodostumisessa. Simon (1956) havainnollistaa yksilö-ympäristö-vuorovaikutusta esimerkillä yksinkertaisesta organismista, jota ohjaa yksi ainoa tarve – ruoka – ja joka on kykeneväinen kolmeen eri toimintoon: lepäämiseen, tutkimiseen ja ruoan hakemiseen. Organismin ympäristö on maasto, jossa on siellä täällä yhtä ateriaa vastaavia ruokamättäitä. Organismi kykenee tarkkailemaan maastoa tietyn säteisellä ympyrällä olinpaikastaan, pystyy liikkumaan rajoitetun matkan, kuluttaa energiaa tietyllä vakio tahdilla ja pystyy syömään rajoitetun määrän kerrallaan. Tällöin rationaalisena pidettävässä käyttäytymisessä organismi tutkii ympäristöä sattumanvaraisesti etsien ruokamätästä ja sellaisen löytäessään kävelee sen luokse ja syö sen. Jos yhden annoksen etsimiseen ja sen luokse siirtymiseen käytettävä energia on keskimäärin pienempi, kuin yhden ruoka-annoksen antama energia, voi organismi käyttää lopun aikansa lepäämiseen. (Simon 1956) Simonin (1956) esittämä kuvaus organismin rationaalisesta päätöksenteosta eroaa taloustieteellisestä, optimointiin pyrkivästä päätöksenteon näkemyksestä. Simon (1956) listaa neljä organismin päätöksentekoa yksinkertaistavaa tekijää:

1) Organismilla on yksi tavoite: ruoka. Sen ei tarvitse miettiä vaihtoehtoisia tavoitteita tai laskea hyötyfunktioita ja piirtää kuvaajia valitakseen toimintamallinsa.
2) Organismin tavoitetaso on vakio, sen ei tarvitse maksimoida ruoan saantiaan. Organismin tulee ylläpitää tietty ruokailutahti, mutta ylimääräisestä ruoasta ei ole organismille hyötyä.
3) Organismin rajallinen etäisyys havainnoida ympäristöään rajoittaa sen suunnitteluhorisonttia. Koska ruokamättäät sijaitsevat ympäristössä sattumanvaraisesti, ruoan etsinnälle ei ole tarvetta laatia säännönmukaista mallia.
4) Organismin tarpeet ja ympäristö luovat luonnollisen jaon keinojen ja päämäärien välille. Ruokamättäitä lukuun ottamatta kaikki ympäristön pisteet ovat organismille yhtä mieleisiä. Tällöin liikkuminen on tarkoituksenmukaista vain ruoan haun tarkoituksessa.

Simon (1956, 131) lisää, että kolme viimeisintä ovat tekijöistä merkittävimmät. Niin kauan kuin tavoitetaso on tilanteen suhteen vakio, suunnitteluhorisontti rajallinen ja keinojen ja päämäärien raja selvä, ei useamman tavoitteen välillä tasapainottelu aiheuta päätöstilanteessa juurikaan epäselvyyttä.

2.2 Päätöksenteon vinoumat

Simonin (1955) rajoitetun rationaalisuuden käsitettä mukaillen Tversky ja Kahneman (1974, 1124) toteavat ihmisen päätösten perustuvan usein arvatuille todennäköisyyksille epävarmoista tapahtumista. Todennäköisyyksien arviointiin käytetään heuristisia yksinkertaistuksia, joilla alun perin monimutkaisempi tilanne saadaan näyttämään selkeämmältä. Yksinkertaistusten käyttö voi olla tiedostettua tai tiedostamatonta (Tversky ja Kahneman 1974, 1124). Tversky ja Kahneman (1974) myöntävät, että heuristiset arviot ovat välillä varsin toimiva työkalu, mutta saattavat usein johtaa vakavaan systemaattiseen virheeseen. Tversky ja Kahneman (1974) esittelevät kolme heuristista kokonaisuutta, joita käytetään usein tilanteiden yksinkertaistamiseen mutta joihin kaikkiin liittyy systemaattisen virheen riski: edustavuus, saatavuus ja ankkurointi.

2.2.1 Edustavuus

Päätöksenteon taustalla on usein jokin seuraavista todennäköisyyden arviointia edellyttävistä kysymyksistä. (1) Millä todennäköisyydellä A kuuluu luokkaan B? (2) Mikä on todennäköisyys, että tapahtuma A syntyy prosessin B seurauksena? (3) Millä todennäköisyydellä prosessista B seuraa tapahtuma A? Edellä esitetyissä kysymyksissä päätöksentekijä nojaa usein edustavuusheuristiikkaan, jossa todennäköisyyttä arvioidaan A:n edustuksena B:stä. Mikäli A:n edustus joukossa B on korkea, arvioidaan todennäköisyys myös korkeaksi. Mikäli edustus on matala, arvioidaan todennäköisyys vastaavasti matalaksi. Tämä johtaa siihen, että seikat, joilla ei ole vaikutusta edustavuuteen, mutta joilla on vaikutus todennäköisyyteen, jätetään helposti tarkastelun ulkopuolelle. Tällaisia ovat esimerkiksi otoksen koko tai joukon osuus perusjoukosta. (Tversky ja Kahneman 1974)

Tversky ja Kahneman (1974, 1125) havainnollistavat edustavuusheuristiikkaa tutkimuksellaan, jossa koehenkilöille näytettiin kirjoitettuja kuvauksia henkilöistä, joiden kerrottiin kuuluvan 100:sta henkilöstä koostuvaan otokseen insinöörejä ja lakimiehiä. Koehenkilöitä pyydettiin arvioimaan millä todennäköisyydellä heidän lukemansa kuvaus kuuluisi insinöörille tai lakimiehelle. Eri koeryhmille annettiin eri lähtötiedot insinöörien ja lakimiesten osuuksista 100 hengen otoksessa. Ensimmäiselle ryhmälle kerrottiin otoksen koostuvan 70:stä insinööristä ja 30:stä lakimiehestä, toiselle ryhmälle ilmoitettiin otoksessa olevan 30 insinööriä ja 70 lakimiestä. Annettujen lähtötietojen perusteella eri koeryhmien kuvauksista antamat arviot tulisivat siis poiketa selvästi toisistaan. Näin ei kuitenkaan käynyt vaan ryhmät arvioivat todennäköisyydet identtisesti. Mielenkiintoinen havainto oli myös, että mikäli kuvauksessa tarjottu informaatio ei antanut viitteitä henkilön kuuluvan kumpaankaan insinöörien tai lakimiesten ryhmään, unohdettiin lähtötiedoissa annettu otosta kuvaileva tieto ja arvioitiin hänet 0,5:n todennäköisyydellä insinööriksi. Mikäli koeryhmille kerrottiin vain insinöörien ja lakimiesten osuus otoksesta, ilman muita lisätietoja, osattiin todennäköisyys kuitenkin arvioida ongelmitta näiden perusteella. (Tversky ja Kahneman 1974, 1124-1125)

2.2.2 Saavutettavuus

Ihmiset ovat taipuvaisia arvioimaan tietyn tapahtuman ilmenemistiheyttä sen perusteella, kuinka helposti he voivat tuoda yhden tapauksen mieleen. Yleisesti tämä on hyödyllistä,

sillä saavutettavuus johtaa siihen, että suuren joukon tapahtumat muistetaan helpommin, kuin pienen joukon. Saavutettavuuteen vaikuttavia tekijöitä on kuitenkin muita kuin pelkkä todennäköisyys, mistä seuraa helposti systemaattisesti toistuvia vinoumia. Muun muassa tietyn tapahtuman tuttuus, vaikuttavuus tai kognitiivinen helppous vaikuttavat saavutettavuuteen, mutta eivät tapahtuman todennäköisyyteen. (Tversky ja Kahneman 1974, 1127)

Joukot joiden esimerkkialkio on tutumpi, on helpompi palauttaa mieleen ja ne arvioidaan usein todennäköisyydeltään suuremmiksi. Tversky ja Kahneman (1974, 1127) havainnollistavat tätä tutkimuksella, jossa koehenkilöille esitettiin lista, jossa oli sekä tunnettujen naisten, että miesten nimiä. Koehenkilöitä pyydettiin tämän jälkeen arvioimaan, oliko listalla enemmän naisia vaiko miehiä. Listat, joissa mainitut naisten nimet olivat mainittuja miesten nimiä kuuluisampia, arvioitiin sisältävän enemmän naisia, kun taas listat, joissa miesten nimet olivat tunnetumpia, arvioitiin olevan enemmän miehiä. (Tversky ja Kahneman 1974, 1127)

Vaikuttavat, emotionaalisesti aktivoivat tapahtumat arvioidaan myös usein todellista todennäköisimmiksi. Esimerkiksi tulipalon nähnyt henkilö arvioi tulipalot yleisimmiksi kuin vain tulipaloista lukenut henkilö. Myös tapahtumien joukot, joiden alkion keksiminen vaatii enemmän kognitiivista ponnistelua, arvioidaan usein todennäköisyydeltään pienemmäksi, kuin tapahtumat, joiden esimerkki on kognitiivisesti helpompi rakentaa, esimerkiksi sanoja jotka alkavat r-kirjaimella arvioidaan helposti olevan enemmän kuin sanoja, joissa r on kolmantena kirjaimena. (Tversky ja Kahneman 1974, 1127)

2.2.3 Ankkurointi

Yksilöt tekevät arvion usein aloittamalla tietystä lähtöarvosta, jota sitten hienosäädetään päätyen lopulliseen arvioon. Eri lähtöarvo johtaa usein erilaiseen loppuarvioon, joka on vinoutunut lähtöarvon suuntaan. Tällaista lähtöarvoon pohjautuvaa päättelyä kutsutaan ankkuroinniksi (Tversky ja Kahneman 1974, 1128)

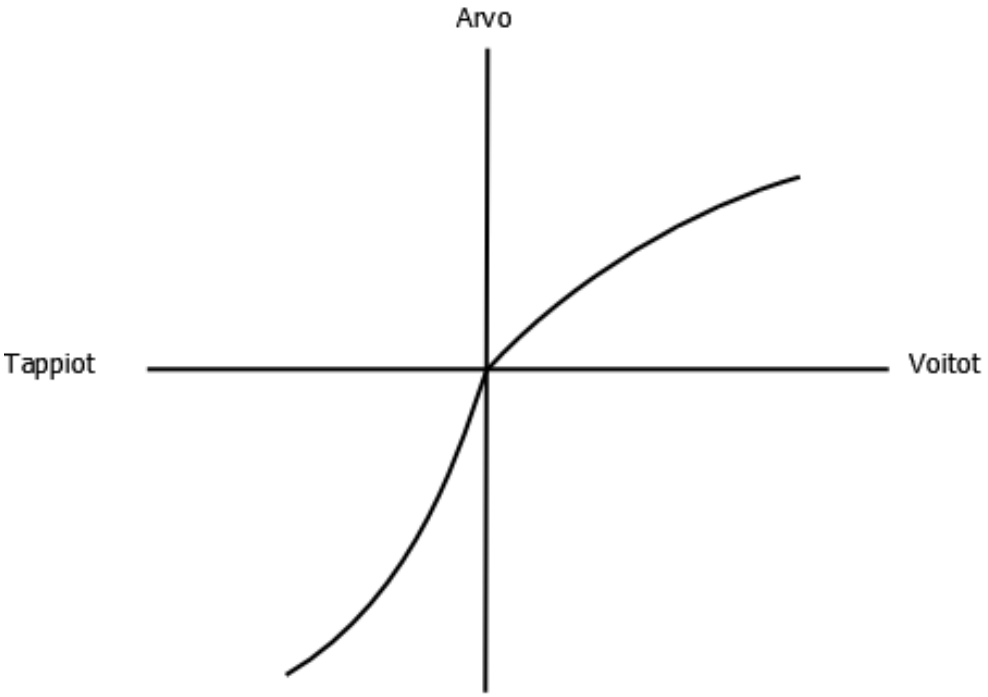
Tversky ja Kahneman (1974, 1128) esittelevät koetta, jossa ryhmää pyydettiin arvioimaan, kuinka monta prosenttia Afrikan maista kuuluu YK:hon. Ennen vastauksen antamista ryhmän edessä pyöräytettiin onnenpyörää, jonka sektorit olivat numeroitu ykkösestä sataan. Ennen lopullisen arvion antamista ryhmältä kysyttiin olisiko lopullinen

arvio yli, vai alle onnenpyörän osoittaman luvun. Ryhmien, joiden pyörät pysähtyivät kohtiin 10 ja 65 keskimääräiset arviot olivat 25 ja 45 prosenttia. (Tversky ja Kahneman 1974, 1128)

Ankkurointi-ilmiö esiintyy usein myös monivaiheisten prosessien tai riskien arvioinnissa. Esimerkiksi uuden tuotteen kehittämisen onnistuminen arvioidaan usein yläkanttiin, sillä arvio ankkuroidaan yksittäisen vaiheen onnistumisen todennäköisyyteen. Vaiheiden lukumäärän lisääntyessä onnistumisen kokonaistodennäköisyys kuitenkin laskee. Kompleksisten kokonaisuuksien tai laitteiden vikariski taas usein aliarvioidaan sillä yksittäisen osan tai komponentin toimintavarmuus on suuri. (Tversky ja Kahneman 1974, 1128)

2.2.4 Prospektiteoria

Edustavuuden, saavutettavuuden ja ankkuroinnin lisäksi yksilön päätöksentekoon vaikuttavat myös päätöksen epävarmuus ja aiemmin koettu menestys. Prospektiteoria kuvaa inhimillistä päätöksentekoa riskiä sisältävissä, epävarmoissa tilanteissa. Teoria pohjautuu havaintoon, etteivät yksilöt useinkaan arvioi esimerkiksi kokonaisvarallisuuteen vaikuttavia päätöksiään kokonaisvarallisuuden muutoksen vaan yksittäisen status quo -pisteen kautta. Teorian mukaan koettava arvo ei ole suoraan verrannollinen päätöksestä mahdollisesti seuraavaan reaaliseen hyötyyn tai tappioon. Kuviossa yksi piirretty voittojen ja tappioiden arvofunktio on kovera valitun status quo -pisteen yläpuolelta ja kupera sen alapuolelta. Tämä tarkoittaa, että esimerkiksi saavutetun hyödyn ero 50€:n ja 150€:n voitoissa koetaan suuremmaksi kuin 1050€:n ja 1150€:n voitoissa. Tappioiden puolella samaten saavutetun hyödyn menetys 50€:n ja 150€:n välillä mielletään suuremmaksi kuin 1050€:n ja 1150€:n välillä. Funktion kuvaaja on myös jyrkempi tappioiden, kuin voittojen puolella. Teorian mukaan tappiosta koituva negatiivinen arvo koetaan suurempana, kuin saman suuruisesta voitosta koituva positiivinen arvo. (Kahneman ja Tversky 1979, 1981, 1982, 1986)



Kuvio 1: Koettu arvo suhteessa voittoihin tai tappioihin (Mukaillen Kahnmen ja Tversky 1984, 342)

Koska päätöksestä koituvia mahdollisia positiivisia tai negatiivisia seurauksia suhteutetaan tiettyyn neutraaliin pisteeseen, nousee pisteen valinnan merkitys kriittiseksi sillä voitot yhdessä pisteessä saattavat näyttäytyä tappioina toisessa pisteessä. Kahneman (1981, 456) havainnollistaa tätä raviesimerkillä, jossa ennen päivän viimeistä lähtöä henkilö on jo hävinnyt 190$. Viimeisessä lähdössä 20:1 veto 10$ panoksella nousee tarkasteluun. Tilanteessa tarkastelupiste voidaan valita kahdella tavalla. Vedonlyöjä voi tarkastella vetoa nykyisen tilanteensa kautta todennäköisenä mahdollisuutena menettää 10$. Todennäköisempi vaihtoehto kuitenkin on, että vedonlyöjä valitsee tarkastelupisteeksi tilanteen ennen päivän aikana syntyneitä tappioita, jolloin tilanne nähdään mahdollisuutena palata valittuun tarkastelupisteeseen. (Kahneman 1981, 456)

2.3 Epärationaalinen sitoutuminen

Epärationaalisella sitoutumisella tarkoitetaan yksilön, ryhmän tai organisaation pitäytymistä tietyssä toimintamallissa, vaikka ulkoiset indikaattorit osoittaisivat suunnan muuttamisen tarpeelliseksi. Kirjallisuus identifioi neljä epärationaalista sitoutumista selittävää kokonaisuutta: psykologisten tekijöiden kokonaisuus, sosiaalisten tekijöiden kokonaisuus, projektitekijöiden kokonaisuus sekä rakenteellisten tekijöiden kokonaisuus.

2.3.1 Psykologiset tekijät

Psykologisilla tekijöillä viitataan päätöstä tekevän yksilön henkilökohtaisiin psykologisiin taipumuksiin. Tästä näkökulmasta kirjallisuus on selittänyt epärationaalista sitoutumista itseoikeutuksen ja prospektiteorian kautta. (Arkes ja Blumer 1985; Staw 1976; 1981; Staw ja Fox 1977; Staw ja Ross 1978; Whyte 1986). Itseoikeutus nähdään kirjallisuudessa dominanttina epärationaalista sitoutumista psykologisesta näkökulmasta selittävänä tekijänä (Staw 1976; 1981; Staw ja Fox 1977; Staw ja Ross 1978). Selityksen taustalla on oletus kompetenssivetoisesti motivoituvasta yksilöstä, jolla on tarve todistaa itselleen ja muille olevansa pätevä ja rationaalinen (Staw 1981). Tällainen egodefensiivisyys johtaa tilanteeseen, jossa yksilö pyrkii suojautumaan psykologisesti rationalisoimalla negatiiviseen lopputulokseen johtaneita toimia (Whyte 1986, 311). Menneiden virheiden rationalisoinnin ja sitä kautta nykyisen toiminnan oikeuttamisen on huomattu entisestään lisäävän panostuksia negatiivisia tuloksia tuottavaan suuntaan ja näin altistavan uusille virheille (Whyte 1986, 311). Päätöksentekijän itseoikeutuksen tarpeen on myös havaittu aiheuttavan negatiivisen palautteen huomiotta jättämistä tai todellisen tilanteen kaunistelua (Staw 1976). Lisäksi on korostettu, että yksilöt ovat taipuvaisia sitomaan enemmän resursseja, mikäli he ovat olleet itse vastuussa negatiivisista seurauksista (Staw 1976).

Staw ja Ross (1978) esittelevät prospektiivisen ja retrospektiivisen rationaalisuuden käsitteet. Prospektiivisella rationaalisuudella kuvataan klassisen rationaalisuuden kaltaista ilmiötä, jossa harkinta kohdistetaan tulevaisuudessa odotettavissa oleviin voittoihin ja tappioihin. Retrospektiivisessä rationaalisuudessa huomioidaan tulevien hyötyjen ja kustannusten lisäksi myös menneisyydessä koituneet tappiot tai saavutetut voitot (Staw ja Ross 1978). Stawin (1980) mukaan päätöksentekijän egodefensiivisyys määrittää kumpi rationaalisuuden laji esiintyy dominanttina yksilön päätöksenteossa. Henkilön vastuu negatiivisista lopputuloksista määrittelee egodefensiivisyyttä. Mikäli henkilö kokee olevansa henkilökohtaisesti vastuussa negatiivisista lopputuloksista, on egodefensiivisyys korkealla ja retrospektiivinen rationaalisuus esiintyy dominanttina päätöksenteossa. Mikäli henkilö ei koe vastuuta negatiivista tuloksista tai saatu palaute on ollut positiivista, egodefensiivisyys on alhaisempaa ja prospektiivinen rationaalisuus esiintyy dominanttina. (Staw 1980) Egodefensiivisyyden ja dominantin rationaalisuuden suhde johtaa siihen, että ”tappioputkissa” päätöksenteko on taipuvaista sisällyttämään uponneet kustannukset harkintaan.

Arkes ja Blumer (1985) tutkivat yksilöiden halukkuutta jatkaa investoimista hankkeeseen, jossa ei ollut taloudellisesti järkevää lopputulosta. Tutkimuksessa todettiin henkilöiden jatkavan investointeja, vaikka he eivät olleet vastuussa uponneista kustannuksista. Vastaavaa käyttäytymistä on havaittavissa myös uhkapeleissä, joissa tappioista on helppo syyttää huonoa tuuria (Whyte 1986, 114). Löydökset eivät ole linjassa Stawin (1976) itseoikeutusteorian kanssa, jonka mukaan vastuun ulkoistamisen pitäisi johtaa alhaiseen egodefensiivisyyteen ja prospektiiviseen rationaalisuuteen. Whyte (1986) näkee Kahnemanin ja Tverskyn (1979) prospektiteorian vaihtoehtoisena selittävänä tekijänä epärationaaliselle sitoutumiselle.

Whyten mukaan (1986, 316) juuri tarkastelupisteen valinta toimii selittävänä tekijänä epärationaaliselle sitoutumiselle. Whyten (1986, 316) mukaan tilannetta tarkastellaan jo kertyneiden voittojen tai tappioiden erona, mikä koetun arvon käyrän muodon vuoksi johtaa riskihakuisuuteen tappion aikana ja riskin välttelyyn voittojen aikana. Toisin kuten Stawin (1976) itseoikeutusteoriassa, prospektiteoriassa päätöksentekijä ei huomioi uponneita kustannuksia suoraan päätöksentekoprosessissa, vaan ne vaikuttavat tarkastelupisteeseen, jonka suhteen päätöksen mahdollisia lopputuloksia punnitaan. Sekä itseoikeutuksen, että prospektiteorian tapauksessa uponneiden kustannusten vaikutus päätöksentekijään on saman suuntainen mutta selittävä mekanismi on erilainen.

2.3.2 Sosiaaliset tekijät

Brockner ja Rubin (1985) eivät näe, että epärationaalinen sitoutuminen olisi selitettävissä puhtaasti yksilöiden psykologisten ominaisuuksien kautta. Päätöksenteon mahdollisesti sosiaalinen luonne vaikuttaa myös epärationaalisen sitoutumisen riskiin. Ryhmän osallistuminen päätöksentekoon ei automaattisesti vaikuta epärationaaliseen sitoutumiseen myönteisesti tai kielteisesti, mutta ryhmässä tapahtuvan kommunikaation on havaittu vahvistavan taipumusta, joka on ryhmälle luontainen (Brockner ja Rubin 1985, 54). Ryhmät, joiden toimet ennakoivat epärationaalista sitoutumista ilmensivät näitä toimia enemmän ryhmätapaamisen jälkeen, kuin ennen tapaamista. Vastaavasti ryhmät, jotka eivät näyttäneet taipumusta epärationaaliselle sitoutumiselle näyttivät sitä tapaamisen jälkeen yhä vähemmän. (Whyte 1990)

Drummondin (1994, 593) mukaan ulkopuolisten todistajien läsnäolo nostaa päätöksentekoa ympäröivää sosiaalista painetta suurentaen näin riskiä epärationaaliseen

sitoutumiseen. Tällainen sosiaalinen paine sisältää muun muassa tarpeen kasvojen säilyttämiseen (Brockner, Rubin ja Langi, 1981), halun kilpailuun tai kostoon (Teger, 1980), epävarmuuden työstä (Fox ja Staw 1979), säännönmukaisesta käyttäytymisestä palkitsemisen (Staw ja Ross 1980), ja henkilökohtaisen identifioitumisen hankkeeseen (Staw ja Ross 1987).

2.3.3 Projektitekijät

Psykologisten ja sosiaalisten tekijöiden ohella myös päätöksenteon kohteena olevan projektin luonne vaikuttaa sekä koettuihin päätöksissä pitäytymisen aiheuttamiin hyötyihin, että tapaan käsitellä uponneita kustannuksia. Erityisesti pitkään kestävien hankkeiden tapauksessa projektiin uponneiden kustannusten on havaittu aiheuttavan epärationaalista sitoutumista. (Brockner, Rubin ja Lang 1981) Myös hankkeesta vetäytymisen kustannukset ovat merkittävä projektikohtainen epärationaalista sitoutumista selittävä tekijä. Erityisen merkittävää tämä on hankkeissa, joiden pelastamisen kustannukset ovat hillittyjä suhteessa vetäytymisen kustannuksiin. Northcraft ja Wolf (1984) käyttävät esimerkkinä osittain asennettua kaasuputkea, jonka poistaminen kustannukset nousisivat lähes yhtä korkeiksi kuin hankkeen viimeistelyn. Myös vaihtoehtoisten investointivaihtoehtojen vähäisyys saattaa joissain tilanteissa selittää epärationaalista sitoutumista (Bateman 1983).

Bowenin (1987) mukaan psykologiset ja sosiaaliset tekijät eivät myöskään pysty selittämään epärationaalista sitoutumista tilanteissa, joissa päätöksessä pitäytyvä henkilö ei ole saanut negatiivista palautetta aiemmista päätöksistään. Tällöin epärationaalinen sitoutuminen saattaa Bowenin (1987) mukaan seuraavista tekijöistä: (a) päätöksen tehneellä henkilöllä oli oletus, että tehdyt päätökset olivat taloudellisesti kannattavia, (b) päätöksen tekijä kokee uuden päätöstilanteen mahdollisuudeksi saada strategia toimivaan, (c) päätöksentekijällä on tarve kokeilla mikäli panosten kasvattaminen auttaa saamaan projektin eloon, tai (d) päätöksentekijä haluaa kerätä dataa ja lisätä ymmärrystään tilanteesta.

2.3.4 Rakenteelliset tekijät

Staw ja Ross (1987, 60) huomauttavat, että monilla rakenteellisilla tekijöillä, jotka eivät suoraan koske käsiteltävää asiaa on vaikutusta myös laajemmalti. He mainitsevat

esimerkiksi työpaikan vaihdon, jolloin päätöstä tekevä henkilö ei ainoastaan vertaa eroja kahden työpaikan välillä. Päätökseen vaikuttaa mahdollinen asuinpaikan vaihto, lasten koulun järjestäminen ja sosiaalisten suhteiden muuttuminen. Rakenteelliset seikat saattavat ajan myötä myös vähentää päätöksenteollista liikkumisvapautta. Staw ja Ross (1987, 60) käyttävät esimerkkinä uuden tuotteen kehittämistä. Tuotteen valmistaminen saattaa edellyttää uutta tuotantolaitosta, laitehankintoja ja uuden henkilöstön rekrytoimista. Tuotteeseen tehdyt rakenteelliset sitoumukset asettavat rajoitteen yrityksen suunnitteluhorisontille tuotepäätöksen jälkeen.

2.4 Kaksi systeemiä

Kahneman (2011) popularisoi tutkimustyönsä kirjassaan Thinking Fast and Slow. Kirjassa hän esittelee ajatuksen kahdesta systeemistä, joista kumpikin lähestyy ongelmanratkaisua ja päätöksentekoa omalla tavallaan. Kahneman (2011) havainnollistaa systeemeitä kahdella eri ongelmalla. Ensimmäisessä ongelmanratkaisijan tehtävänä on katsoa alla olevaa kuvaa.

Kuvio 2: Tunnetilan intuitiivinen tunnistaminen (Kahneman 2012, 19)

Kuviosta kaksi havaitsemme välittömästi, ilman kognitiivisia ponnisteluja, että kuvattu nainen on vihainen. Sen lisäksi todennäköisesti ennustamme tietynlaista käyttäytymistä, mahdollisesti kovaan ääneen lausuttuja epämiellyttäviä ilmaisuita. Sekä tunnetilan tunnistaminen että mahdollisen tulevan käyttäytymisen ennakointi tapahtuivat automaattisesti, ilman tietoista ponnistelua. Tällaista tiedostamatonta ja nopeaa johtopäätösten tekemistä Kahneman (2011, 20) kutsuu nopeaksi ajatteluksi.

Toisessa tilanteessa ongelmanratkaisijan tulee ratkaista seuraava laskutoimitus:

$$17 * 24$$

Tunnistamme tehtävän välittömästi kertolaskuongelmaksi ja todennäköisesti pystymme karkeasti rajaamaan lukualuetta, jolle vastaus todennäköisesti asettuisi. Suurin osa meistä ei kuitenkaan pysty ensimmäisen esimerkin tapaan intuitiivisesti vastaamaan esimerkiksi siihen onko vastaus 468 vaiko ei. Pystymme myös päättämään ohjaammeko kognitiivista kapasiteettiamme ongelman ratkaisuun vaiko luovutammeko yrittämättä. Tällaista toisen esimerkin kaltaista tilannetta Kahneman (2011, 20) nimittää hitaaksi ajatteluksi. Hitaalle ajattelulle on tyypillistä tietoinen kognitiivinen työskentely ongelman ratkaisemiseksi. Esimerkin kertolaskutehtävässä ongelmanratkaisijan tulee ensiksi noutaa mielestään koulussa opittu, ketolaskun suorittamiseen tarkoitettu kognitiivinen ohjelma. Ohjelman suorittaminen vaatii tiedon väliaikaista varastoimista muistiin. Ratkaistakseen esitetyn laskutoimituksen ongelmanratkaisijan tulee pysyä kartalla siitä missä on tällä hetkellä menossa ja mikä on ohjelman seuraava askel.

Kahneman (2011) korostaa kahden edellä mainitun systeemin erilaisuutta toisiinsa verrattuna. Systeemi 1 toimii autonomisesti ja nopeasti, vähäisellä tai täysin olemattomalla ponnistelulla ja ilman kontrollointia. Se vastaa toiminnoista kuten: objektien etäisyyksien vertailu, vihamielisyyden havaitseminen äänestä, yllättävän äänen suunnan tunnistamisesta, yksinkertaisten lauseiden ymmärtämisestä ja auton ajamisesta tyhjällä tiellä. Kahneman (2011) liittää myös intuition käsitteen systeemin 1 ominaisuudeksi.

Siinä missä systeemi 1 toimii autonomisesti, intuitiivisesti ja yksilön tietoisesta valvonnasta irrallaan. Systeemi 2 allokoi tietoisesti huomiota kognitiivista ponnistelua vaativiin tehtäviin, kuten esimerkissä mainittuun laskutoimitukseen. Toisin kuin systeemi 1, systeemi 2 yhdistetään tietoiseen toimintaan, valintaan ja keskittymiseen. Se vastaa toiminnoista kuten: sosiaalisesti suotavan käyttäytymisen monitorointi, puhelinnumeron kertominen, veroilmoituksen täyttäminen tai loogisen argumentin validiuden arviointi. (Kahneman 2011, 20-22)

Kahneman (2011) liittää inhimillisen päätöksenteon vinoumat systeemin 1 piirteeksi. Näin siitäkin huolimatta, että vinoumat ilmenevät usein kognitiivisesti haastavimmissa tehtävissä, joiden ratkaisu tapahtuu pääsääntöisesti systeemin 2 toimesta. Systeemin 2 toimintaan vaikuttavat systeemin 1, ilman tietoista päätöstä, assosiatiivisesti muistista noutamat faktat ja ehdotukset (Kahneman 2011).

2.5 Asiantuntijuus päätöksenteossa

Asiantuntija on henkilö, jonka tunnustetaan kykenevän suoriutumaan tehtävästä korkeimmalla mahdollisella tasolla (Shanteau 1992, 255). Kirjallisuus (Chase ja Simon 1973; Kahneman ja Klein 2009; Shanteau 1992) liittää asiantuntijuuteen myös intuition käsitteen, kyvyn tehdä asiantuntijuuden aluetta koskevia päätöksiä nopeasti ja suurella tarkkuudella. Kirjallisuus lähestyy asiantuntijan intuitiota kahdesta näkökulmasta. Yhtäältä asiantuntijat nähdään omalla alueellaan luontaisina ja harjaantuneina päätöksentekijöinä (DeGroot 1978; Chase ja Simon 1973), ja tutkimus keskittyy selittämään syitä, jotka mahdollistavat asiantuntijan nopean, intuitiivisen päätöksenteon. Tällaista näkökulmaa kutsutaan luontaisen päätöksenteon näkökulmaksi (*eng. Naturalistic Decision Making)* (DeGroot 1978; Chase ja Simon 1973). Toisaalta asiantuntijoita pidetään erehtyväisinä, päätöksenteon vinoumille alttiina ja asiantuntija-asemansa ansiosta yliluottavaisina toimijoina (Meehl 1954; Tversky ja Kahneman 1971). Tällöin tutkimuksen keskiössä on asiantuntijan virhealttiuden selittäminen. Asiantuntijuuteen kriittisemmin suhtautuvaa näkemystä kutsutaan heuristiikoiden ja vinoumien näkökulmaksi *(eng. Heuristics and Biases)*. Edellisessä kappaleessa esitetyn systeemiajattelun näkökulmasta asiantuntijan intuition kehitys voidaan nähdä kokonaisuuksien siirtymisenä tietoista kognitiivista työskentelyä edellyttävästä systeemin kaksi alaisuudesta nopeamman systeemi yhden alle.

Luontaisen päätöksenteon näkökulman juuret voidaan katsoa löytyvän Chasen ja Simonin (1973) ja DeGrootin (1978) shakin pelaajille tehdyistä tutkimuksista. DeGroot (1978) osoitti, että shakkimestarit näkivät parhaat siirrot intuitiivisesti ja nopeasti siinä missä keskivertopelaajat eivät edes harkinneet näitä siirtoja. DeGroot (1978) vahvisti Chasen ja Simonin (1973) aiemmat löydökset. Chase ja Simon (1973) kuvasivat shakkimestarien taitoa kyvyksi tunnistaa monimutkaisia kuvioita ja säännönmukaisuuksia ja arvioivat, että parhaimpien pelaajien repertuaariin kuuluisi noin 50 000 – 100 000 erilaista pelilaudalta välittömästi tunnistettavaa tilannetta. Näiden tilanteiden alitajuinen tunnistaminen mahdollisti siirron löytämisen intuitiivisesti, ilman tarvetta analysoida suurempaa joukkoa mahdollisia siirtovaihtoehtoja. Työnsä pohjalta Chase ja Simon (1973) määrittelivät intuition kyvyksi tunnistaa muistiin varastoituja säännönmukaisuuksia.

Meehl (1954) nosti esiin asiantuntijuuteen kriittisemmin suhtautuvan heuristiikkojen ja vinoumien näkökulman. Meehl (1954) kävi läpi noin 20 tutkimusta, joissa verrattiin asiantuntijan, useimmissa tapauksissa kliinisen psykologin suoriutumista ennakointitehtävässä yksinkertaista tilastollista mallia vastaan. Ennustettavat tapahtumat vaihtelivat akateemisen menestyksen ennakoinnista väkivallan ja rikoksen uusiutumisen ennustamiseen. Vaikka tilastollisissa malleissa käytettiin vain osaa asiantuntijoiden saatavilla olevasta informaatiosta, olivat yksinkertaiset tilastolliset mallit silti asiantuntijoita parempia ennustajia lähes kaikissa tapauksissa. Meehl (1954) uskoi, että asiantuntijoiden heikko suoriutuminen johtuu jonkinlaisesta systemaattisesta virheestä, kuten lähtökohtaisten todennäköisyyksien huomioimattomuudesta yksittäistä tapausta tarkasteltaessa. Muun muassa Kahneman ja Tversky (1979; 1981; 1982; 1984), Simon (1955; 1956), Staw (1976; 1981) Staw ja Fox (1977) Staw ja Ross (1978) sekä Whyte (1986) ovat myöhemmin tarkemmin identifioineet päätöksentekoon ja päätöksissä pitäytymiseen vaikuttavia tekijöitä, jotka ovat läsnä myös asiantuntijan päätöksenteossa. Kyseiset tekijät on tarkemmin esitelty tämän tutkimusraportin kappaleissa 2.2 ja 2.3. Kahneman ja Klein (2009) mainitsemat myös epäsäännönmukaisuuden erääksi suurimmaksi informaalin päätöksentekotilanteen heikkoudeksi. Eri asiantuntijat päätyvät saman lähtöaineiston perusteella usein eri lopputuloksiin.

Kahneman (2003) kuvailee asiantuntijan kokemaa valheellista tunnetta päätöksen validiteetista omien kokemustensa kautta. Toimiessaan Israelin armeijan psykologisen tutkimuksen yksikössä eräs hänen tehtävistään oli arvioida upseerikoulutukseen pyrkiviä kokelaita. Kahneman (2003) kuvailee vahvaa jokaiseen kandidaattiin tutustumisen tunnetta, jonka perusteella hän uskoi olevansa kykeneväinen ennustamaan kuinka kyseinen kandidaatti pärjäisi myöhemmässä koulutuksessa ja taistelutilanteessa. Subjektiivinen tunne kyvystä tulkita jokainen tapaus yksilöllisesti ei vähentynyt, vaikka tilastollinen palaute upseerikoulutuksesta osoitti, että valintamenettelyn luotettavuus oli vähintäänkin kyseenalainen (Kahneman 2003). Tämä Kahnemanin (2003) kuvailema ilmiö on linjassa Einhornin ja Hogartin (1978) tekemän havainnon kanssa. Heidän mukaansa yksilöt eivät kykene itse arvioimaan, onko onnistuneen päätöksen pohjalla taitoa vaiko pelkkää yliluottamusta. Myöskään päätöksentekijän subjektiivisen arvion ja päätöksen onnistumisen välillä ei ole löydettävissä merkittävää korrelaatiota (Einhorn ja Hogart 1978).

Sekä luontaisen päätöksenteon että heuristiikkojen ja vinoumien näkökulma määrittelevät intuition Chasea ja Simonia (1973) mukaillen kyvyksi tunnistaa päätöksentekoympäristöstä nousevia ärsykkeitä. Molemmat näkökulmat myös tunnistavat, että asiantuntijan intuitio voi näyttäytyä toisissa tilanteissa loisteliaana ja toisissa haparoivana ja yliluottavaisena. Koulukunnat jakavat myös käsityksen asiantuntijan intuitiivisen päätöksenteon kuuluvan kappaleessa 2.4 käsitellyn systeemin 1 ominaisuudeksi. Asiantuntijan intuitiivinen päätöksenteko on usein automaattista ja ympäristön tarjoamat ärsykkeet nousevat mieleen vaivattomasti ilman ponnisteluja ja tietoista harkintaa. Luontaisen päätöksenteon ja heuristiikkojen ja vinoumien näkökulmat kuitenkin eroavat tavassa lähestyä intuitiota. Luontainen päätöksenteko keskittyy opettelun ja harjaantumisen kautta tapahtuvaan intuitiiviseen käyttäytymiseen, kun taas heuristiikat ja vinoumat tarkastelevat heuristisen yksinkertaistuksen kautta tapahtuvaa intuitiota.

Kahneman ja Klein (2009) hahmottelevat artikkelissaan reunaehtoja ympäristölle, jossa asiantuntijan intuition kehittyminen on mahdollista. Heidän mukaansa päätöksentekoympäristön tulee tarjota vakaa suhde objektiivisesti havaittavien ärsykkeiden ja ärsykkeitä seuraavien tapahtumien, tai ärsykkeiden ja niiden pohjalta suoritettujen toimintojen lopputulosten välillä. Edellä mainittuja suhteita Kahneman ja Klein (2009) kuvaavat ympäristön validiteetin käsitteellä. Esimerkiksi lääketieteen tai palontorjunnan voidaan katsoa tarjoavan edellä mainittujen kriteerien valossa suhteellisen korkean validiteetin päätöksentekoympäristön intuition kehittymiselle. Vastaavasti yksittäisen osakkeen kurssin ennustaminen tai pitkän aikavälin ennuste poliittisista tapahtumista tehdään matalan validiteetin ympäristössä. (Kahneman ja Klein 2009, 523-525) Kahneman ja Klein (2009, 524) korostavat, etteivät päätösympäristön validiteetti ja epävarmuus ole toisiaan poissulkevia, vaan ympäristö voi samanaikaisesti olla sekä epävarma että validi. Kahneman ja Klein (2009, 524) mainitsevat pokerin ja sodankäynnin esimerkeiksi tämänkaltaisista ympäristöistä.

Korkean validiteetin ympäristö on edellytyksenä asiantuntijan intuition kehittymiselle. Tämän lisäksi yksilön tulee kyetä oppimaan ympäristön säännönmukaisuuksia. Käytännössä tämä edellyttää nopeaa ja yksiselitteistä palautetta ympäristöltä. (Kahneman ja Klein 2009, 524) Kahneman ja Klein (2009, 525) huomauttavat, että vaikka intuitiivisen taidon oppiminen vaihtelevissa ja epävarmoissa ympäristöissä on mahdollista, tekevät yksilöt toisinaan päätöksiä, joiden lopputulokset ovat puhtaasti

sattumanvaraisia. Tällaisissa tilanteissa taidon illuusion ja yliluottamuksen kehittyminen on todennäköistä. Varsinkin kun ihmiset eivät kykene erottelemaan sattuman ja taidon osuutta tiettyyn lopputulokseen pääsemisessä (Einhorn ja Hogart 1978).

2.6 Vahvat ja toimivat heuristiikat

Gigerenzer (2008) sekä Gigerenzer ja Brighton (2009) kritisoivat vahvasti Kahnemanin työtä heuristisen päätöksenteon vinoumien parissa. Gigerenzer (2008) huomauttaa, etteivät edustavuus, saavutettavuus tai ankkurointi juurikaan tarjoa toimivaa rajausta ja ilman selkeää päätöksentekoa kuvaavaa mallia yleismaailmalliset määrittelyt eivät ole tarpeeksi rajaavia. Hän myös kritisoi Kahnemanin (2011) kahden systeemin ajattelua kohtuuttomasta yleistämisestä. Gigerenzer (2008) korostaa, ettei heuristiikkoja voi ylimalkaisesti lakaista systeemi 1 otsikon alle ja yhdistää päätöksenteon vinoumiin. Gigerenzer (2008, 21) listaakin kuusi yleistä heuristiikkoihin liittyvää väärinymmärrystä, jotka on esitetty taulukossa yksi.

Taulukko 1: Heuristiikkoihin liittyvät väärinymmärrykset (Gigerenzer 2008, 21)

Väärinymmärrys	**Tarkennus**
1. Heuristiikat ovat aina kakkosvaihtoehto, optimointi on joka tilanteessa parempi.	Monissa tilanteissa optimointi on laskennallisesti mahdotonta tai epätarkempaa arviointivirheen vuoksi.
2. Luotamme heuristiikkoihin vain kognitiivisten rajoitteidemme vuoksi.	Käytämme heuristiikkoja mielen ja ympäristön piirteiden yhteisvaikutuksen vuoksi.
3. Heuristiikkoihin luotetaan vain pienen merkityksen rutiinipäätöksissä.	Heuristiikkoihin luotetaan sekä suuren, että pienen merkityksen päätöksissä.
4. Suuremman kognitiivisen kapasiteetin omaavat ihmiset käyttävät tilastollisia menetelmiä, pienemmän kapasiteetin omaavat heuristiikkoja.	Ei empiirisiä todisteita.
5. Edustavuus, saavutettavuus ja ankkurointi ovat heuristisia malleja.	Käsitteet ovat pelkkiä nimikkeitä, eivät formaaleja malleja. Mallin tulee olla ennustettava ja empiirisesti testattava.
6. Enemmän informaatiota ja laskentaa johtaa aina parempaan päätökseen.	Usein päätöksen tekeminen osittain epävarmassa ympäristössä edellyttää informaation huomioimattomuutta.

Gigerenzer ja Brighton (2009) korostavat erityisesti, että vastoin yleistä käsitystä laskennan, käsiteltävän informaation ja käsittelyajan lisääminen ei välttämättä nosta arvioiden tarkkuutta. Heuristiikkojen tutkimus on päinvastoin osoittanut, että säästö

laskennassa, informaatiossa ja ajankäytössä voi vaikuttaa tarkkuuteen positiivisesti (Gigerenzer ja Brighton 2009, 108). Heuristiikkoja käytetäänkin erityisesti tilanteissa, joissa päätös tulee tehdä nopeasti ja informaatiota on rajallisesti tai se on ristiriitaista (Gigerenzer 2008, 20). Laskennan ja käsiteltävän informaation suhdetta lopputulokseen käsitellään tarkemmin tutkimusraportin kappaleessa 4.5, jossa käsitellään heuristiikkojen suoriutumista matemaattisia malleja vastaan.

Gigerenzer (2008) kuvaa ihmismieltä sopeutuvana työkalupakkina, joka koostuu useista eri tilanteissa käytettävistä heuristiikoista. Gigerenzer ja Brighton (2009, 130) kokoavat yhteen 10 empiirisesti tuettua heuristiikkaa, mutta korostavat, ettei listaa voida pitää mitenkään valmiina tai täysin kattavana. Heuristiikat on listattu taulukkoon kaksi.

Taulukko 2: Empiirisesti tuetut heuristiikat

Heuristiikka	**Määritelmä**	**Ekologisesti rationaalinen jos**	**Empiiriset löydöt**
Tunnistamis-heuristiikka (Goldstein ja Gigerenzer 2002)	Jos toinen kahdesta vaihtoehdosta tunnistetaan, määritetään sille korkeampi valintakriteerin arvo	Tunnistamisen validiteetti > 0.5	
Luontevuus-heuristiikka (Jacoby ja Dallas 1981)	Mikäli molemmat vaihtoehdot tunnistetaan, mutta toinen nopeammin, määritetään nopeammin tunnistetulle korkeampi valintakriteerin arvo	Luontevuuden validiteetti > 0.5	
Valitse paras (Gigerenzer ja Goldstein 1996)	1) Käy läpi tilanteeseen liittyviä vihjeitä validiteettijärjestyksessä. 2) Lopeta läpikäynti, kun vihje antaa viitteitä valinnalle 3) Valitse vaihtoehto, jota vihje suosii		Saavuttanut regressioanalyysiä (Czerlinski ym. 1999), neuroverkkoja ja päätöspuualgoritmeja (Brighton, 2006) korkeamman ennustavuuden.
Vakiokertoimet (Dawes 1979)	Valintaa tehtäessä arvotetaan kaikkia positiivisia vihjeitä vakiokertoimella	Eri vihjeiden kelpoisuus vaihtelee vähän	Useissa kokeissa saavuttanut regressioanalyysiä suuremman ennustavuuden. (Czerlinski ym. 1999)

Tyydyttävyys (Simon 1955)	Valitaan vaihtoehdoista ensimmäinen, joka täyttää tavoitetason	Vaihtoehtojen määrä vähenee ajan kuluessa	
1/*N* Tasajako-heuristiikka (DeMiguel 2009)	Allokoidaan resurssi tasan kaikkien *N* vaihtoehtojen kesken	Pieni ennustettavuus, pieni opetusjoukko, suuri *N*.	
Oletusarvo-heuristiikka (Johnson ja Goldstein 2004)	Mikäli on olemassa oletusarvoinen vaihtoehto, valitaan se.	Oletusarvon asettajien arvot täsmäävät päätöksentekijän arvojen kanssa ja seurauksia on vaikea ennakoida.	
Tit-for-tat (Axelrod 1984)	Pyri ensin yhteistyöhön ja matki sitten toisen osapuolen viimeisintä suoritetta	Tilanne sallii kyseisen heuristiikan käytön ja toinen osapuoli käyttää samaa heuristiikkaa.	
Matki enemmistöä (Boyd ja Richerson 2005)	Havainnoi enemmistöä ja matki heidän käyttäytymistään	Ympäristö on vakaa tai vaihtuu hitaasti, tiedonhaku työlästä tai aikaa vievää.	Ajuri ryhmän identiteetin ja moraalisen normiston rakentumisessa.
Matki menestyneintä (Boyd ja Richerson 2005)	Valitse näkemyksesi mukaan menestynein henkilö ja matki hänen käyttäytymistään.	Oppiminen hidasta, tiedonhaku työlästä tai aikaa vievää	Kulttuurisen evoluution ajuri

Esitetyistä heuristiikoista Gigerenzerin ja Goldsteinin (1996) valitse paras -heuristiikka on Gigerenzerin (2008; Goldstein ja Gigerenzer 2002; Gigerenzer ja Brighton 2009) eniten viittaama. Valitse paras -heuristiikkassa päätös tehdään vain yhden muuttujan perusteella ja jätetään muut muuttujat huomioimatta, vaikka niiden arvo olisi tiedossa. Muistista noudetaan valintakriteeriä ennakoivia vihjeitä järjestyksessä merkityksellisimmästä vähemmän merkitykselliseen. Heti kun löytyy muuttuja, jonka suhteen vaihtoehdot eroavat, valitaan vaihtoehto tämän muuttujan perusteella ja jätetään muut huomiotta. (Gigerenzer ja Brighton, 1996)

Simonin (1955; 1956) tavoin, myös Gigerenzer (2008) korostaa, ettei päättely tapahdu koskaan irrallaan ympäristöstä. Huomion tulisikin keskittyä määrittämään kussakin tilanteessa ja ympäristössä parhaiten toimiva päätöksentekostrategia, koostui se sitten

todennäköisyyksistä, logiikasta, heuristiikoista tai niiden yhdistelmistä (Gigerenzer 2008, 21). Gigerenzer (2008) kuvaa oikean päätösstrategian valintaa ekologisen rationaalisuuden käsitteellä. Gigerenzer (2008, 25) muistuttaa, ettei Darvinistisesta lähtökohdasta organismin tavoite ole olla rationaalinen, vaan toimia tavoitetasonsa ohjaamana. Toisin kuin taloustieteellinen logiikkaan ja todennäköisyyksiin pohjautuva rationaalisuuden määritelmä, ekologinen rationaalisuus määrittyy suhteessa tapahtumaympäristöön, eikä suhteessa optimiin. Gigerenzer ja Brighton (2009, 129) huomauttavat, ettei, varsinkin heuristisen strategioiden tapauksessa, pääosin alitajuista strategianvalinnanprosessia tunneta tarkoin. Kirjallisuudessa on kuitenkaan identifioitu neljä päätösstrategian valintaan vaikuttavaa tekijää.

Ensimmäinen valintaan vaikuttava tekijä on muistin asettama rajoite. Muistista palautettavissa oleva informaatio rajoittaa suoraan yksilön käytössä olevien heuristiikkojen määrää. (Gigerenzer ja Brighton 2009, 129) Jos henkilön tulee valita kahdesta vaihtoehdosta valintakriteerit paremmin täyttävä, mutta hän tunnistaa vaihtoehdoista vain toisen eikä tunne muuta valintaan liittyvää informaatiota, rajoittuu käytettävä päätösstrategia tunnistamis-heuristiikkaan. Mikäli henkilö tuntee molemmat vaihtoehdot, sekä muistaa vaihtoehtoihin liittyvää informaatiota, rajoittuu tunnistamis-heuristiikka pois ja vaihtoehdoiksi jäävät luontevuus-heuristiikka ja valitse-paras-heuristiikka. Tehdyissä kokeissa suurin osa henkilöistä siirtyi käyttämään tietopohjaista heuristiikkaa (kuten valitse paras), mikäli molemmat vaihtoehdot tunnetaan ja päätöksen kannalta merkittävää informaatiota muistetaan (Gigerenzer ja Brighton 2009).

Saatu palaute ja oppiminen on toinen päätösstrategian valintaan vaikuttava tekijä. Rieskamp ja Otto (2006) esittelevät strategian valinta oppimisen, missä ympäristöstä saatu palaute muokkaa kyseisessä ympäristössä käytettävien päätösstrategioiden preferenssijärjestystä. Kyse on vahvistusoppimisesta, jossa tietyn käyttäytymismallin sijaan ympäristön palaute vahvistaa tai heikentää tietyn päätösstrategian käyttöpreferenssiä jatkossa. (Rieskamp ja Otto 2006)

Kustannus ja saavutettu hyöty on myös nähty laajalti päätösstrategian valintaan vaikuttavana tekijänä (Beach ja Mitchell 1978, Christensen-Szalanski 1978, Payne ym.1993). Kustannuksella viitataan päätöksen vaatimaan kognitiiviseen panostukseen ja hyödyllä strategian tarkkuuteen. Tapoja hyötyjen ja kustannusten määrittelemiseen on kuitenkin erilaisia. Beach ja Mitchell (1978) korostavat päätöksentekijän

henkilökohtaisia ominaisuuksia ja preferenssejä hyötyjen ja kustannusten määrittämisen tavoissa.

Viimeiseksi päätösstrategian valintaan vaikuttavaksi tekijäksi Gigerenzer ja Brighton (2009, 132) listaavat tehtäväympäristön rakenteen. Tällä he tarkoittavat ympäristöstä saadun vihjeen validiteettia suhteessa valintakriteeriin. Validiteetilla viitataan vihjeen kykyyn ennakoida valintakriteeriä. Esimerkiksi kaupungin nimen tunnistaminen on validiteetiltaan korkea vihje, mikäli valitaan vaihtoehdoista suurinta kaupunkia. Mikäli valittavana on kaupunki, joka on lähinnä rannikkoa, tunnistamisvihjeen validiteetti on alhaisempi. (Gigerenzer ja Brighton 2009 129 - 131)

2.7 Päätöksenteko organisaatiossa

Useista päätöksentekijöistä koostuvilla kokonaisuuksilla ei välttämättä ole samanlaista yhtenäistä päätöksentekoa ohjaavaa tavoitetilaa kuin yksittäisellä päätöksentekijällä (Simon 1979). Marchin (1988) mukaan huomion allokointi, konfliktit sekä säännöt ja rituaalit ovat erityisesti organisaatioiden päätöksentekoon liittyviä kokonaisuuksia mutta jakavat yhteisiä piirteitä myös yksilön päätöksenteon kanssa.

2.7.1 Huomion allokointi

Organisaatiossa tarkasteluun nostettavien vaihtoehtojen valinta ohjaa lopullista päätöstä enemmän kuin vaihtoehtojen välinen vertailu (March 1988). Tätä ilmiötä March (1988) kutsuu huomion allokoinniksi. Aivan kuten yksilön tarkasteluhorisontti, myöskin organisaatioiden kyky suunnitella toimintaansa eri vaihtoehtojen pohjalta on rajallinen. March ja Simon (1958) korostavat, että organisaatiot eivät kykene päätöstilanteissa selvittämään kaikkia mahdollisia vaihtoehtoja, tai niiden valitsemisesta aiheutuvia seurauksia. Lisäksi organisaatiot eivät käytännössä pysty ajamaan kaikkia asettamiaan tavoitteita samanaikaisesti (Cyert ja March 1963). Koska organisaatiot kykenevät käsittelemään vain rajattua määrää vaihtoehtoja, seurauksia ja tavoitteita, ohjaa päätöksentekoa enemmän tarkastelun painopiste, kuin vaihtoehtojen objektiivinen vertailu. (March 1988, 3) Tätä huomion allokoinnista seuraavaa tarkastelun painopisteen siirtämistä March (1988, 3) kutsuu organisatoriseksi etsinnäksi.

Organisatorista etsintää ohjaa Marchin (1988, 3) mukaan kaksi tekijää. (1) Organisaatiot näkevät menestyksen ja epäonnistumisen yleensä hyvin mustavalkoisena ilmiönä, eikä

menestyksen tai epäonnistumisen asteita tunnisteta. Organisaatiot ovat taipuvaisia ohjaamaan enemmän huomiota toimintoihin, jotka ovat epäonnistuneet tavoitteissaan. (2) Menestyksen aikana organisatorisen etsinnän määrä vähenee, epäonnistumisen aikana lisääntyy. (March 1988, 3) Organisaatioiden etsintäponnistelut siis määräytyvät koetun epäonnistumisen funktiona. Organisaatorisen etsinnän voidaan todeta noudattavan kappaleessa 2.2.4 esiteltyä Kahnemanin (1984, 342) prospektiteorian arvofunktiota, jonka mukaan tappion aikana koettu menetys koetaan itseisarvoltaan suurempana kuin onnistumis-putkessa koettu saman suuruinen voitto. Organisaatiot myös pyrkivät suojautumaan epäonnistumiselta ohjaamalla enemmän etsintäponnisteluja riskiä sisältävien vaihtoehtojen tullessa vastaan (March ja Shapira, 1987). Tällöin energiaa käytetään enemmän kilpailevien vaihtoehtojen etsimiseen (March ja Shapira, 1987). Etsinnän lopputuloksena organisaatio valitsee vaihtoehdon, jolla se olettaa pääsevänsä asetettuun tavoitteeseen. Vaihtoehdot eivät siis kilpaile keskenään, vaan niitä suhteutetaan aina lopputavoitteeseen.

2.7.2 Konfliktit

Useista yksilöistä ja ryhmistä koostuvat organisaatiot sisältävät usein sisäisesti ristiriitaisia tavoitteita ja prioriteettijärjestyksiä. Aikaisimpia taloustieteellisiä näkemyksiä lukuun ottomatta kirjallisuus tunnustaa organisaatiot poliittisia piirteitä omaaviksi järjestelmiksi, jotka eivät usinkaan tee päätöksiä kaikkien jakamien tavoitteiden pohjalta (Cyert ja March 1959, March 1962). Organisaation rakenteiden sisällä yksittäiset jäsenet tai ryhmät saattavat käyttää organisaation resursseja, kuten informaatiota, vipuna omien tavoitteidensa ajamiseen. Hillitäkseen tällaista käyttäytymistä organisaatioihin muodostuu käytänteitä ja rakenteita, jotka tasapainottavat mahdollisia yksilöiden välisiä näkemyseroja tavoitteissa ja priorisoinnissa. Eräs esimerkki tällaisesta käytänteestä ovat muun muassa työsopimukset, joissa määritellään yksilöä koskevat vastuualueet ja tehtävän rajaukset. (March 1988, 8)

March (1988, 9) tunnistaa kolme huomion rajallisuuteen liittyvää konfliktien leviämistä hillitsevää tekijää. Organisaation hierarkia kaventaa tietyn päätösprosessin yleisöä, jolloin kaikki mahdollisesti omien tavoitteiden kanssa ristiriitaiset tekijät eivät ole näkyvissä kaikille organisaation jäsenille yhtäaikaisesti. Organisaatiossa toimivien yksilöiden käytettävissä oleva aika ja energia ovat rajallisia. Tämän vuoksi yksilöiden huomio rajautuu vain oman tehtävän kannalta kriittisimpiin päätöksiin ja päätösajureihin.

Myös organisaation käytössä mahdollisesti olevat ylimääräiset resurssit antavat liikkumavaraa asetettujen tavoitteiden ja vaadittavan suoritustason vaihtelun mukaan. (March 1988, 9) Konfliktien leviämistä hillitsevien tekijöiden vuoksi risteävistä tavoitteista johtuvat konfliktit leviävät organisaatiossa ennemminkin jaksottaisesti kuin yhtäaikaisesti. Käytänne, joka korjaa tilanteen yhdessä osastossa, aiheuttaa ongelman toisessa, jonka korjaaminen todennäköisesti heijastuu jälleen organisaation muihin osiin. (Cyert ja March, 1963)

2.7.3 Säännöt ja käytänteet

Useissa tilanteissa organisaation päätöksentekoa määrittää enemmän säännöt ja käytänteet kuin päätökseen liittyvien vaihtoehtojen ja niistä koituvien seurausten vertailu (March ja Simon 1958). Organisaatioilla on yleensä päätöksentekoon liittyviä formaaleita tai vähemmän formaaleita toimintamalleja, jotka ohjaavat yksilöiden käyttäytymistä ja joilla on vaikutus päätöksen lopputulokseen. Marchin (1981) mukaan varsinkin usein toistuvissa säännöllisissä päätöksissä on huomattu, että tiettyjä sääntöjä ja käytänteitä noudatetaan ennemminkin siksi, että niitä on opittu pitämään tietyissä tilanteissa sopivina tai tiettyyn tehtävään kuuluvina, kuin siksi, että ne auttaisivat itse päätöksenteossa.

Kirjallisuuden voidaan katsoa jakavan näkemys siitä, että säännöt ja rutiinit ovat organisaation historian muovaamia. Ne yhtäältä heijastelevat, mutta toisaalta eivät tallenna aiemman toiminnan kautta opittua tietoa (March 1981). Historiasta johdettujen käytänteiden parissa tehty tutkimus (March ja Olsen 1975) osoittaa, että käytänteillä on taipumus olettaa yhtäläisyyttä tilanteiden, jossa käytänne on luotu ja tilanteiden, jossa sitä sovelletaan välillä. Tämä johtaa tilanteeseen, jossa ympäristössä tapahtuvia muutoksia ei enää pystytä havaitsemaan.

2.8 Havaintojen rikastaminen

Sekä yksittäiset ihmiset, että organisaatiot kykenevät tietyissä tilanteissa rikastamaan päätöksenteossa hyödynnettävää ympäristöstä saatavaa informaatiota. March ym. (1991) antavat viisi esimerkkiä erilaisista tilanteista, joissa organisaatiot eivät ole läpikäyneet tapahtumia, mutta joissa historia, tarjoaa työkaluja tuottaa päätöksenteossa hyödynnettävää tietoa. Esimerkiksi (1) sotaorganisaatiot osallistuvat vain harvoin taisteluun mutta pyrkivät silti kehittämään kykyään sodankäynnissä. (2) Yrityksillä ei

usein ole kokemusta kansainvälisistä investoinneista mutta pyrkimys ja kyky tarkastella investoinnin onnistumista historiansa kautta. (3) Lentoyhtiöillä on harvoin tuhoisia onnettomuuksia. Siitä huolimatta historiatietoja tarkastelemalla pyritään oppimaan ja pienentämään sellaisten todennäköisyyttä. (4) Radikaalit innovaatiot ovat harvinaisia. Yritykset kuitenkin tarkastelevat toimintaansa, jotta voivat oppia lisäämään tuottamiensa innovaatioiden lukumäärää. (5) Ydinonnettomuuksia tapahtuu harvoin. Samoin kuin lentoyhtiöt, energian tuottajat pyrkivät minimoimaan tuhoisien onnettomuuksien mahdollisuuden. March ym. (1991) esittävät kaksi mekanismia, joiden kautta organisaatiot pystyvät oppimaan yllä esitetyn kaltaisissa tilanteissa: tapahtumien rikas kokeminen ja kokemuksen simulointi.

Tapahtumien rikas kokeminen viittaa toimintaan, jossa organisaatiot käsittelevät tapahtumia yksityiskohtaisina narratiiveina. Marchin ym. (1991, 1) mukaan historian tarkastelu yksittäisinä datapisteinä aliarvioi sen tarjoamia oppimisen mahdollisuuksia. Yksittäiset datapisteet, kuten päätöksestä koituva lopputulos, tulee toki ottaa tarkasteluun. Ennen lopputuloksen koitumista organisaatio kuitenkin kokee päätöksenteosta itsestään aiheutuneita seurauksia, jotka tarjoavat mahdollisuuksia oppimiseen ennen päätöksen seurausten realisoitumista. Tällaisia päätöksentekoa ympäröiviä seikkoja ovat esimerkiksi ilmapiiri päätöksentekohetkellä ja sen jälkeen, tai päätökseen liitetyt mielikuvat kuten ”rohkea liike” tai ”hyvä kompromissi”. Mikäli päätöksen tekemisestä koituvat aikaiset kokemukset ovat positiivisia, yksilöt ovat usein taipuvaisia vahvistamaan saman suuntaista toimintaa. March ym. (1991, 2) korostavat, että tällaisten ympäröivien tekijöiden vaikutukset päätöstä arvioitaessa voivat merkitä enemmän kuin lopulta koituva lopputulos.

Organisaatiot myös usein keskittyvät intensiivisesti tarkastelemaan merkittävinä pidettyjä tapahtumia. March ym. (1991, 2) listaavat kolme tekijää, jotka vaikuttavat tapahtuman merkityksellisyyteen. Historiallisen kehityksen risteyskohdat, jotka muuttavat maailmaa kriittisesti ovat merkittäviä. Tällaisten tapahtumien intensiivisestä tarkastelusta opitaan usein enemmänkin tulevaisuuden muuttuneita implikaatioita kuin sitä, miten ennustaa tai kontrolloida vastaavia tapahtumia vastaisuudessa. March ym. (1991, 2) käyttävät kirjapainon keksimistä esimerkkinä tällaisista tapahtumista. Kehityksen risteyskohtien lisäksi, myös tapahtumat, jotka muuttavat sitä mitä maailmasta uskotaan, ovat merkittäviä. Yksittäiset tapahtumat harvoin ovat yllättäviä, sillä ne mahtuvat vallitsevan teorian sallimaan vaihteluväliin. Toisinaan yksittäinen tapahtuma kuitenkin nostaa

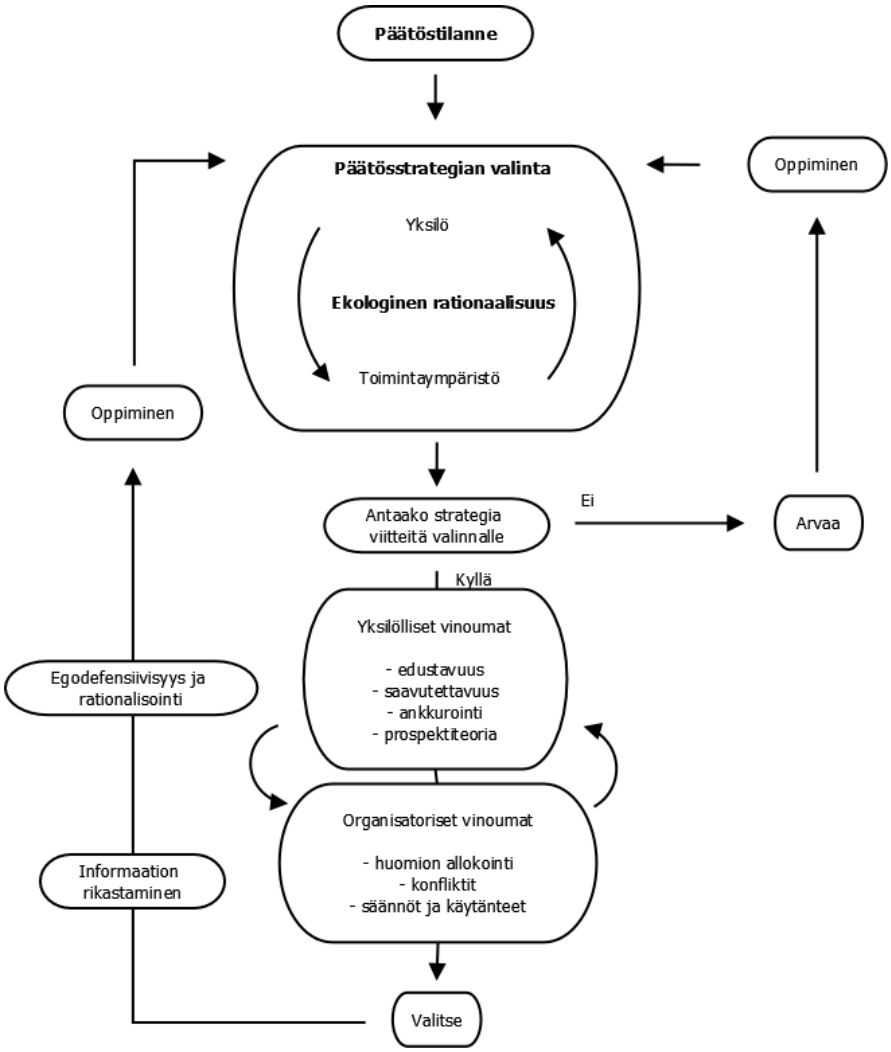
nykyisten uskomusten kanssa ristiriidassa olevia seikkoja ja kyseenalaistaa aikaisemmin totena pidettyjä uskomuksia. Uutta mielenkiintoa ja merkitystä organisaation sisällä synnyttävät metaforallisesti voimakkaat tapahtuvat ovat myös merkittäviä. Tarinankerronnallinen kyvykkyys on usein merkittävä voima tällaisten tapahtumien tapauksessa. March ym. (1991, 2) kuitenkin korostavat, että myös myös materialliset seikat ja raaka kokemus voivat synnyttää metaforallisesti merkittäviä tarinoita.

Organisaatiot kykenevät tehostamaan merkittävistä tapahtumista oppimista sekä tulkitsemalla niitä lukuisista eri näkökulmista että arvottamalla tapahtumiin liittyvää onnistumista ja epäonnistumista erilaisilla mittareilla. Organisaatioissa tällainen toiminta tapahtuu usein rutiininomaisesti virallisten käytänteiden sekoittuessa epäformaaleihin keskusteluihin ja organisaatiossa eläviin tarinoihin. Näin syntyy yhteisesti tulkittava historia. (March ym. 1991, 2)

Yksittäistä tapahtumaa voidaan tarkastella mahdollisena lopputulemana laajassa mahdollisten tapahtumien ympäristössä. Tällöin tapahtumien rikkaan kokemisen lisäksi organisaatiot kykenevät simuloimaan kokemusta luomalla kuvauksia erilaisista mahdollisista olosuhteista, jotka olisivat voineet vallita yksittäisen tapahtuman hetkellä ja näin saada monipuolisempaa tietoa yksittäisestä tapahtumasta. (March ym. 1991,3) March ym. (1991, 3) käyttävät lentoliikennejärjestelmiä esimerkkinä kokemuksen simuloinnista. Järjestelmän toimijat keräävät tietoa läheltä piti tilanteista piloteilta ja lennonjohtajilta. Tämän tiedon pohjalta muodostettujen skenaarioiden avulla he kehittävät koneita, lennonjohdon järjestelmiä ja lentäjien koulutusta.

2.9 Päätöksenteon ja intuition yhteen kietoutuminen

Prosessinomaisesta lähtökohdasta kuvattuna inhimillistä päätöksentekoa havainnollistava runko rakentuu päätösstrategian valinnan sekä strategian käytöstä seuraavien mahdollisten vinoumien ympärille. Kuviossa kolme havainnollistetaan inhimillisen päätöksenteon rakennetta päätöstilanteen ilmenemisestä lähtevänä kaaviona. Päätösstrategian valinta voi tilanteesta riippuen olla joko tietoista, systeemin 2 ohjaamaa, tai tiedostamatonta, systeemin 1 intuitiivista toimintaa.



Kuvio 3: Inhimillisen päätöksenteon prosessikuvaus

Gigerenzerin (2008) kuvailema ekologinen rationaalisuus nähdään mallissa dominantiksi päätösstrategian valintaan vaikuttavaksi tekijäksi. Käsitteellisesti tämä on hyvin lähellä Simonin (1955; 1956; 1957) huomioita yksilön pyrkimyksestä ympäristövuorovaikutuksessa muodostettavaan tyydyttyneisyyden tasoon. Ekologinen rationaalisuus suhteuttaa samalla tavoin yksilön tavoitetason yksilön toimintaympäristöön. Päätösstrategia valitaan siis aina ympäristön ja yksilön omien tavoitteiden mukaan, ei ympäristön ja optimin mukaan. Yksilön tavoitetason lisäksi päätöksentekijästä riippuvaisia strategian valintaan vaikuttavia tekijöitä ovat muistista palautettavien päätösstrategioiden määrä (Gigerenzer ja Brighton 2009), päätösstrategian käyttämisestä koettu hyöty ja siitä aiheutuva kognitiivinen kustannus (Beach ja Mitchell 1978, Christensen-Szalanski 1978, Payne ym. 1993) sekä aiempien kokemusten kautta kertynyt palaute ja oppiminen (Rieskamp ja Otto 2006). Toimintaympäristö määrittää käyttökelpoisten valintakriteeriä ennakoivien vihjeiden määrän, mikä rajaa mahdollisten

päätösstrategioiden määrää. Vaikka mallissa käytetään päätösstrategian yhteydessä sanaa valinta, on huomioitava, ettei kyse välttämättä ole tietoisesta valinnasta. Varsinkaan jokapäiväisissä päätöksissään yksilöt eivät tietoisesti valitse eri päätösstrategioiden välillä, vaan ympäristöstä saatavan informaation mukaan pikemminkin vain päätyvät käyttämään jotain strategiaa. Toisaalta tämä ei myöskään tarkoita, ettei tietoista päätösstrategian valintaa olisi olemassa. Esimerkiksi asiantuntijarooleissa voidaan käyttää ennalta tiettyä päätöstä varten suunniteltua mallia. Tällöin päätösstrategia ei tyypillisesti ole myöskään puhtaasti heuristinen, vaan todennäköisesti yhdistelmä erilaisia tilastollisia menetelmiä, kriteeristöjä ja heuristiikkoja.

Valittu päätösstrategia on altis sekä kappaleessa 2.2 esitellyille yksilöstä johtuville, että kappaleessa 2.7 listatuille organisatorisille vinoumille. Huomioitavaa on, että sekä yksilön päätöksentekoon liittyvistä, että organisatorisista vinoumista puhuttaessa, käsitellään vinoumia suhteessa optimiin eikä yksilön tai organisaation tavoitetason määrittämään tyydyttyneisyyden tasoon. Vinoumien voidaankin katsoa olevan kriittisimpiä tilanteissa, joissa tietoisesti pyritään hakemaan parasta mahdollista ratkaisua. Reaalimaailmassa yksilölliset ja organisatoriset vinoumat eivät myöskään ole toisistaan irrallisia. Esimerkiksi huomion allokointi tai säännöt ja käytänteet vaikuttavat tiettyjen tapahtumien esiintymiseen organisaation sisällä ja siten päätöksiä tekevien yksilöiden saavutettavuusvinoumaan. Tätä yksilöllisten ja organisatoristen vinoumien vuorovaikutusta kuvataan mallissa kaarevilla nuolilla.

Päätöksestä koituvat seuraukset vaikuttavat oppimisen muodossa sekä yksilöiden että organisaatioiden toimintaan. Kykymme rikastaa tekemiämme havaintoja johtaa siihen, että oppiminen ei tapahdu yksinomaan päätöksestä koituvien seurausten perusteella. Esimerkiksi päätöksestä seuraava tunnetila tai sosiaalinen palaute voivat muodostaa kriittisen osan, kun tulkitsemme tehdyn päätöksen onnistuneisuutta. Lisäksi on huomioitava Kahnemanin ja Kleinin (2009) havainto siitä, ettei toimintaympäristö kaikissa tilanteissa mahdollista oppimisen edellytyksenä olevaa tarpeeksi nopeaa tai säännönmukaista palautetta, Einhornin ja Hogartin (1978) havainto yksilöiden kyvyttömyydestä tunnistaa tuurin ja taidon osuutta saavutetussa lopputuloksessa sekä Stawin (1976; 1980) kuvailema yksilöiden egodefensiivisyydestä johtuva taipumus rationalisoida päätöksiään. Edellä listattujen havaintojen vuoksi oppimisen ei voida katsoa automaattisesti johtavan optimaalisempaan päätösstrategian valintaan jatkossa. Epärationaalinen sitoutuminen voidaankin kuviossa kolme esitetyn mallin mukaan

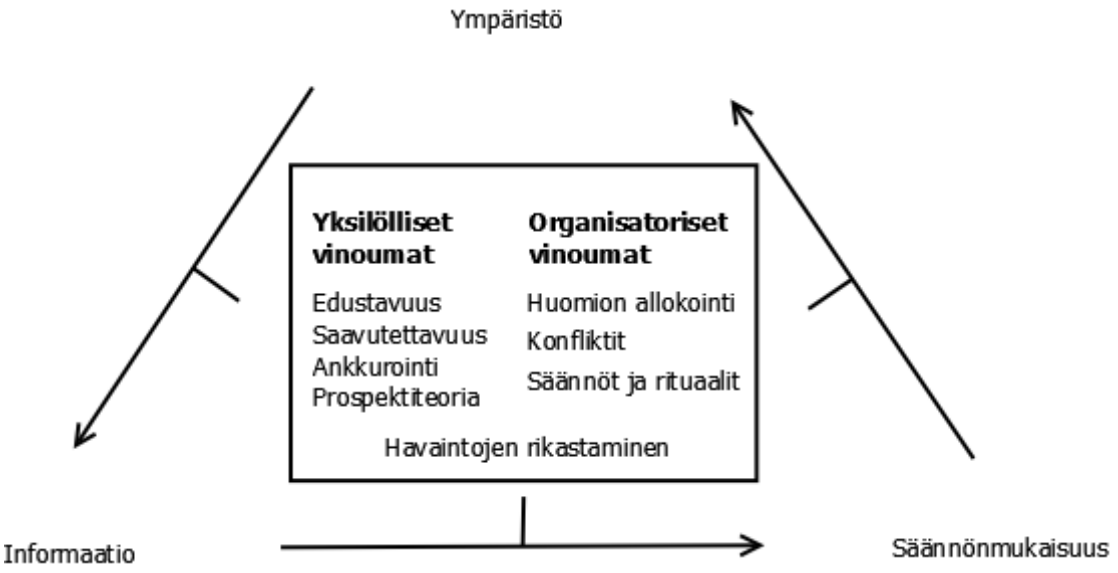
osittain nähdä epäoptimaalisena oppimisena, jossa toimintaympäristön luonne, psykologiset, sosiaaliset, projekti- ja rakennetekijät sekä kyvyttömyys arvioida tuurin osuutta lopputuloksessa johtaa vahvistamaan sekä optimin että yksilön tavoitetason kannalta haitallisten päätösstrategioiden valintaa.

Päätöksenteko voi olla varsinkin usein toistuvissa tilanteissa intuitiivista, systeemin 1 ohjaamaa käyttäytymistä. Lisäksi esimerkiksi kappaleessa 2.5 viitatut shakkimestareille tehdyt tutkimukset osoittavat, että yksilöt ovat kykeneväisiä kehittymään intuitiivisiksi päätöksentekijöiksi omalla asiantuntijuuden alueellaan. Kuviossa kolme havainnollistetussa prosessimaisessa kuvauksessa intuitio näyttäytyy valinnasta tai arvauksesta päätösstrategian valintaan etenevänä, toistuvana oppimisen kehänä. Mikäli sama päätöstilanne toistuu samassa toimintaympäristössä riittävän usein, alkaa päätösstrategian valinta tapahtua suoremmin ja nopeammin toimintaympäristöstä saatavien ärsykkeiden mukaan. Kuviossa kolme esitetyn päätöksenteon prosessikuvauksen lisäksi intuitiota on aiheellista käsitellä myös tarkemmin suhteessa ympäristöön ja päätöksenteon vinoumiin.

Kappaleessa 2.5 käsitelty asiantuntijan intuitio liitetään usein päätöksenteon nopeuteen, oikeellisuuteen ja tarkkuuteen. Koska intuition kehittyminen vaatii säännöllisen palautemekanismin, päätöksestä koituva seuraus muodostaa vain osan palautteesta ja koska yksilöillä on taipumus rationalisoida omaa päätöksentekoa koskevaa informaatiota, on perusteltua väittää, että intuitio voi myös kehittyä tekemään väärän ja epätarkan päätöksen. Erityinen riski tämän tapahtumiseen on, mikäli päätöksen seurauksen ulkopuoliset palautemekanismit ovat toimintaympäristöä säännöllisempiä ja painoarvoltaan seurausta suurempia. Tässä tutkimuksessa intuitio määritelläänkin laajemmin Chasen ja Simonin (1973) mukaan kyvyksi tunnistaa ympäristöstä muistiin varastoituja säännönmukaisuuksia ottamatta kantaa siihen tuottaako muistista palautettu säännönmukaisuus tavoitellun lopputuloksen kannalta suotuisan vai epäsuotuisan lopputuloksen.

Säännönmukaisuuksien muodostumista ja suhdetta ympäristöön sekä päätöksenteon vinoumiin havainnollistetaan kuviossa neljä. Intuition muodostuminen esitetään ympäristön, informaation ja säännönmukaisuuksien välisenä yksisuuntaisena vuorovaikutuksena, jossa ympäristöstä nostettua informaatiota hyödynnetään ensin

säännönmukaisuuden muodostamiseen ja myöhemmin tietyn säännönmukaisuuden muistista palauttamiseen.



Kuvio 4: Säännönmukaisuus suhteessa ympäristöön ja päätöksenteon vinoumiin

Yksilöllisten ja organisatoristen vinoumien sekä havaintojen rikastamisen nähdään moderoivan ympäristön ja siitä nostettavan informaation sekä informaation ja sen avulla muodostetun säännönmukaisuuden suhdetta vaikuttaen näin säännönmukaisuuden kykyyn ennustaa reaalista ympäristöä. Vinoumat ja havaintojen rikastaminen vaikuttavat sekä muistiin varastoituneiden säännönmukaisuuksien muodostumiseen että niiden palauttamiseen. Säännönmukaisuuksien muodostamisen näkökulmasta vinoumat ja havaintojen rikastaminen vaikuttavat sekä huomiomme ohjautumiseen että ympäristöstä nostettavan informaation painoarvoon. Esimerkiksi informaatio, jonka kognitiivinen käsittely on helppoa tai joka sisältää emotionaalista arvoa, todennäköisesti havaitaan nopeammin ja se saa säännönmukaisuuden muodostuessa suuremman painoarvon neutraaliin, hankalammin ymmärrettävään informaatioon verrattuna. Muistiin aiemmin varastoitujen säännönmukaisuuksien palauttamisessa vinoumien ja havaintojen rikastaminen vaikuttaa vastaavalla mekanismilla. Esimerkiksi kogniivinen helppous ja tunnepitoinen tieto voi vaikuttaa myös tietyn säännönmukaisuuden palauttamisen nopeuteen ja vaikuttaa näin meidän tulkintaamme ympäristöstä ja säännönmukaisuudesta, jota oletamme sen seuraavan.

Yllä kuvattu vaikutus informaation painottamiseen saattaa tietyissä tilanteissa johtaa systemaattiseen virheeseen mutta tekee ihmisen päätöksenteosta osaltaan nopeampaa ja mahdollistaa toimimisen vähäisemmällä informaation määrällä. Vinouma sanan käyttäminen inhimillisen päätöksenteon kuvaamisessa on osaltaan harhaanjohtavaa. Vaikka vinoumina tunnettu kokonaisuus onnistuukin kuvaamaan ihmisen tiedonkäsittelyä päätösprosessissa, eivät vinoumat johda automaattisesti erehtymiseen tai epäoptimaalisiin päätöksiin.

3 METODOLOGIA

Tässä luvussa esitellään ja perustellaan tutkimuksen metodologiset valinnat. Luku vastaa tutkimusaineiston keräämistä ja analysointia koskeviin kysymyksiin sekä kuvaa tutkimusprosessin kannalta keskeiset vaiheet. Luvun tehtävänä on asettaa tutkimuksen metodiset valinnat tieteelliselle tarkastelulle alttiiksi ja siten osaltaan parantaa tutkimuksen kokonaisluotettavuutta.

3.1 Laadullinen lähestymistapa

Tämän tutkimuksen tehtävänä on ymmärtää tekoälyjärjestelmien päätöksentekoa ja tehdä siitä vertailtavaa inhimillisen päätöksenteon kanssa. Edellä kuvattu tutkimuksen tavoite on luonteeltaan laadullinen. Tutkimus pyrkii ymmärtämään tekoälyä ilmiönä ja luomaan syvempää tietoa tarkasteltavana olevasta kohteesta päätöksenteon näkökulmasta. Tutkimuskysymyksen luonteen vuoksi tutkimukseen valittiin laadullinen kvalitatiivinen tutkimustapa. Silvermanin (2010, 11) mukaan laadullinen tutkimus on soveltuva juurikin tilanteissa, joissa pyritään vastaamaan kysymykseen ”miten” ennemmin kuin ”kuinka monta”.

Laadullinen tutkimus pohjautuu hermeneuttiseen tiedekäsitykseen, jonka mukaan ihminen eroaa muista organismeista tietoisuutensa, vapaan tahtonsa sekä niille perustuvan kulttuurisen elämänmuotonsa perusteella. Tällöin ihmisen toimintaa ja toiminnan tuloksia – yksilönä ja yhteisönä – tulisi tarkastella tahdottuina ja tarkoitettuina, erilaisia merkityksiä ilmaisevina kokonaisuuksina. Laadulliselle tutkimukselle keskeistä on siten merkitysten tulkinta, ei yleisillä laeilla selittäminen. (Tuomivaara 2005, 29)

Tämän tutkimuksen aineisto koostuu haastatteluista sekä verkkojulkaisuista. Haastattelut suoritettiin puolistrukturoituina, eli teemahaastatteluina. Hyvärisen ym. (2017) mukaan laadullisella menetelmällä suoritettujen tutkimusten haastattelut ovat yleisestikin enemmän tai vähemmän puolistrukturoituja. Teemahaastattelussa haastattelun teemat tiedetään, mutta niiden yksityiskohtainen muoto ja järjestys ovat epäselviä. Teemahaastattelun aikana tutkija ylläpitää keskustelua apukysymysten ja avainsanojen avulla. Tutkimuksen lopputuloksen kannalta on merkittävää, että haastateltavaksi saadaan henkilöitä, joilta tutkija olettaa saavansa tehokkaimmin tietoa tutkittavasta ilmiöstä. (Saarinen-Kauppinen ja Puusniekka 2006, 56)

3.2 Tutkimusprosessin kuvaus

Tutkimusprosessi käynnistyi tutkimusilmiöön tutustumalla. Tekoäly valittiin tarkasteltavaksi kokonaisuudeksi tutkijan oman mielenkiinnon sekä sen odotetun liiketoiminnallisen sekä taloudellisen merkittävyyden vuoksi. Ilmiöön tutustumisen aloitettiin alan tieteellisiin julkaisuihin, seminaariaineistoihin sekä avoimiin verkkojulkaisuihin perehtymällä. Samalla muodostui suurpiirteinen hahmotelma tutkimuksen sekundaariaineistosta. Laine (2018, 34) mainitsee tutkijan esiymmärryksen muodostumisen olevan kriittinen osa tutkimusprosessia ja edellytys merkitysten ymmärtämiselle ja rikkaan tutkimuksellisen dialogin muodostamiselle.

Esiymmärryksen muodostamisen yhteydessä hahmoteltiin mahdollisia teoreettisia tulokulmia tutkimusilmiöön. Päätöksenteko valittiin lopulta käytettäväksi viitekehykseksi, sillä se tarjosi vahvan tiedeyhteisön validoiman kirjallisuuden, eikä aiempi tutkimus ollut käsitellyt tekoälyä päätöksiä tekevänä entiteettinä. Päätöksentekokirjallisuuden katsottiin ankkuroivan tutkimus vahvasti kauppatieteelliseen tutkimusperinteeseen ja samalla luovan uutta, tekoälyn soveltamisessa hyödynnettävää tietoa.

Tutkimusongelma hahmottui yhdessä teorian ja hankitun esiymmärryksen rajautuessa. Ensimmäisten tutkimuskysymysten hahmotelmat pyrkivät tarkastelemaan tekoälyn päätöksentekoa jollain sovellusalueella ja vastaamaan kysymyksiin: millaisissa sovellusaluetta koskevissa päätöksissä tekoälyä voitiin hyödyntää tai millaisia haasteita tekoälyn implementoinnissa olisi. Edellä kuvatut tutkimusongelmat törmäsivät kuitenkin laajuuden hallitsemisen haasteeseen. Tutkimus ei olisi pystynyt säilyttämään tarpeeksi selvää ongelman rajausta ja samalla kuvaamaan sekä tekoälyn päätöksentekoa että sovellusalueen erityispiirteitä tieteellisen uskottavuuden kannalta riittävällä laajuudella. Tekoälyn päätöksenteon ymmärtäminen muodosti kuitenkin reunaehdon myös mahdollisten sovellusaluetta koskevien kysymysten ratkaisemiseksi. Se nähtiin merkittävämpänä kokonaisuutena sovellusalueen erityispiirteisiin verrattuna ja tutkimus päätyi lopulta keskittymään tekoälyn päätöksentekoon.

Yllä kuvattu tutkimusprosessi mukailee hyvin Duboisin ja Gadden (2002) kuvaamaan systemaattista yhdistelyä, jossa tutkija kulkee edestakaisin tutkimusaktiviteettien, empiiristen havaintojen ja teorian välillä. Tutkimusprosessista on jälkikäteen erotettavissa tiettyjä vaiheita, kuten tutkittavan ilmiön, tutkimuskysymysten tai teorian

valinta. Tutkimuksen kuluessa vaiheet eivät kuitenkaan olleet tiukasti lukittuja. Niihin palattiin kiertokulkumaisesti jonkin tutkimusteeman alla olevan kokonaisuuden selventyessä tai uusien kysymysten noustessa.

3.3 Aineiston keruu

Tässä tutkimuksessa käytetty aineisto kerättiin triangulatorisesti, eli monimenetelmällisesti useita eri lähteitä käyttämällä. Eriksson ja Kovalainen (2008) näkevät triangulatorisen lähestymistavan tarjoavan monipuolisemman kuvan tekijöistä ja toiminnoista tietystä ilmiöstä yhteiskunnallisessa ympäristössä. Vaikka tutkimuksen fokus ei korosta yhteiskunnallista ympäristöä, voidaan tutkimuksen kirjallisuuden pohjalta todeta päätöksenteon olevan aiheena vahvasti riippuvainen sovellusympäristöstään. Päätöksenteko jakaa siten yhdistäviä piirteitä yhteiskunnallisen ympäristön kanssa. Tällöin triangulatorista, useaa näkökulmaa tarkastelevaa, prosessia voidaan pitää tutkimuksen tavoitteen näkökulmasta perusteltuna.

3.3.1 Primaariaineisto

Primaariaineistoksi kutsutaan tutkijan itsensä keräämää, välitöntä tietoa tutkimuskohteesta sisältävää aineistoa (Hirsjärvi ym. 2009, 186). Tässä tutkimuksessa primaariaineisto kerättiin puolistrukturoituja teemahaastatteluja käyttämällä. Puolistrukturoidussa teemahaastattelussa on kuudesta kahteentoista kysymystä, jotka on valittu tutkimusongelman ehdoilla. (Rowley, 2012). Tämän tutkimuksen haastatteluissa teemat rakentuivat teknologian kehittämistä koskevien haasteiden ja tekoälyn kuvaamisen ympärille. Käsittelyssä olleet teemat sovitettiin kunkin haastateltavan asiantuntijuuden alueeseen.

Teemahaastattelussa tarkkoja haastattelukysymyksiä ei ole lukittu ennen haastattelun suorittamista. Haastattelu rakentuu valittujen teemojen ympärille keskustelunomaisesti (Hyvärinen ym. 2017, 21). Myöskään haastattelun teemoille ei valittu, aloitusteemaa lukuun ottamatta, etukäteistä käsittelyjärjestystä vaan teemoihin ohjattiin haastattelun aikana keskustelulle luontaisessa järjestyksessä. Ydinteemojen lisäksi haastattelut sisälsivät aloitus ja lopetus osiot, joissa käytiin läpi luottamuksellisuutta ja tietojen käsittelyä koskevat seikat sekä vastattiin haastateltavalla mahdollisesti oleviin, tutkimusta koskeviin kysymyksiin. Ennen jokaista haastattelua haastateltavan kanssa käytiin läpi

haastattelun ja tutkimuksen teemat suurpiirteisellä tarkkuudella. Läpikäynti suoritettiin joko sähköpostitse tai puhelimitse samalla kun haastattelun ajankohdasta sovittiin. Kaikki haastattelut nauhoitettiin myöhempää litterointia ja analysointia varten.

Saarinen-Kauppinen ja Puusniekka (2006, 56) korostavat haastatteluun valittavien henkilöiden merkitystä tutkimuksen lopputulokselle. Tässä tutkimuksessa haastatteluun valinnan edellytyksenä oli asiantuntijuuden tasoiset tiedot tekoälyn kehittämisestä tai sen implementaatioista tietyllä sovellusalueella. Käytännössä tämä edellytti joko toimijuutta alan tutkimuksen tai alaan liittyvän konsultoinnin parissa, tai toimimista tekoälyteknologioita kehittävässä yrityksessä, jolla oli teknologiaan perustuva markkinoilla oleva tuote sekä asiakkaita. Kaikki tässä tutkimuksessa suoritetut haastattelut olivat asiantuntijahaastatteluita.

Asiantuntijahaastattelua ei voida pitää itsenäisenä haastattelumenetelmänä, vaikka sitä ilmentävätkin tietyt erityispiirteet. Asiantuntijalla oletetaan olevan sellaista tietoa tutkittavasta asiasta, jota vain harvalla on. Haastattelijan tavoite on päästä käsiksi juuri tähän tietoon. Tällöin tutkimuksen kohteena ei ole haastateltava henkilö itse vaan tieto tutkimuksen kohteena olevasta aiheesta. (Hyvärinen ym. 2017, 214 – 219) Jokainen haastateltava on vastannut tutkimusaihetta käsitteleviin teemoihin oman subjektiivisen näkemyksensä mukaan, irrallaan edustamastaan organisaatiosta tai nykyisestä työtehtävästään. Tähän tutkimukseen haastatellut henkilöt sekä haastattelun ajankohta ja kesto on listattu taulukossa kolme. Kolme haastateltavaa esittivät toivomuksensa esiintyä haastattelussa nimettöminä. Heihin viitataan tutkimusraportissa nimityksillä asiantuntija sekä heidän edustamaansa organisaatioon nimellä organisaatio A.

Taulukko 3: Tutkimuksen haastattelut, niiden ajankohdat ja kestot

Henkilö	**Päivämäärä**	**Kesto**
Heikki Huttunen	31.5.2018	56min
Reko Lehti	14.11.2018	1h 57min
Asiantuntija 1	15.11.2018	46min
Asiantuntija 2	15.11.2018	1h 07min
Asiantuntija 3	15.11.2018	1h 07min

Heikki Huttunen toimii Tampereen yliopiston signaalinkäsittelyn laboratorion apulaisprofessorina. Hän on kirjoittanut lukuisia koneoppimista ja hahmontunnistusta

käsitteleviä tieteellisiä julkaisuita sekä sijoittunut korkealle alan käytännön osaamista mittaavissa avoimissa kilpailuissa. Tämän lisäksi hän toimii automatisoituja kulunvalvontaratkaisuja tarjoavan Visy Oy:n hallituksessa. Reko Lehti on toiminut 15 vuoden ajan strategia- ja teknologiakonsulttina. Tällä hetkellä hän on partnerina tekoälyyn, teknologioihin ja liiketoiminta-alustoihin erikoistuneessa Taival Advisory:ssä. Asiantuntijat 1, 2 ja 3 toimivat samassa luonnollisen kielen käsittelyyn erikoistuneessa yrityksessä, johon viitataan tässä tutkimusraportissa nimellä organisaatio A. Asiantuntija 1 on eräs yrityksen perustajajäsenistä ja vastaa yrityksen liiketoiminnan kehittämisestä. Asiantuntijat 2 ja 3 toimivat yrityksen datatiimissä ja vastaavat yrityksen tuotteen kehittämisestä. Asiantuntija 2 tekee väitöskirjatutkimusta luonnollisen kielen käsittelystä ja koneoppimisesta. Asiantuntija 3 on koneoppimisesta tutkinnon suorittanut tohtori.

3.3.2 Sekundaariaineisto

Tutkijan itsensä tuottaman primaariaineiston lisäksi tutkimuksessa hyödynnettiin myös muiden koostamaa materiaalia. Tällaista aineistoa kutsutaan sekundaariaineistoksi ja siihen kuuluvat esimerkiksi aiemmissa tutkimuksissa tuotettu materiaali, tilastot, dokumenttiaineistot kuten yritysten omat tiedotteet ja kotisivut (Hirsjärvi, ym. 2009, 186-189) sekä uutiset, selvitykset ja raportit (Koskinen ym. 2005, 131). Primaariaineiston täydentämisen lisäksi sekundaariaineistoa käytettiin tutkimusalueen kartoittamiseen sekä aiheen rajaamiseen. Vaikka huomattava osa aiheen rajaamisessa hyödynnetystä materiaalista jäi lopulta tutkimusraportin ulkopuolelle, oli sillä välillinen vaikutus kirjoittajan näkemyksiin ja siten myös tämän tutkimusraportin lopulliseen ilmeeseen.

Tässä tutkimuksessa primaariaineistoa täydennettiin pääsääntöisesti verkkouutisilla, tiedotteilla sekä raporteilla. Sekundaariaineistoa käytettiin havainnollistamaan primaariaineistosta tehtyjä löydöksiä ja siten sitomaan luonteeltaan käsitteelliseempää primaariaineistoa reaalimaailman tapahtumiin. Tutkimusraportin tekoälyn tekemiä virheitä tarkasteleva viides luku on kirjoitettu lähes kokonaan sekundaariaineiston perusteella. Tutkimuksen sekundaariaineisto koostui avoimista verkkojulkaisuista, joita haettiin käyttämällä yleisesti saatavilla olevia hakukoneita kuten Googlea, Bingiä ja Yahoota. Käytettyjä hakusanoja olivat muun muassa: AI-cases, AI failures, AI failures 2016, AI failures 2017, AI-human interaction ja AI mistakes. Tutkimukseen valikoitujen tekoälyä käsittelevien tapausten oikeellisuudesta on pyritty varmistumaan tarkistamalla uutisen pitävyys vähintään kahdesta eri lähteestä. Facebookin ja Northpointen

tapauksessa asianomaisten yritysten tuottamat tutkimusraportit olivat myös hyödynnettävissä.

3.4 Aineiston analyysi

Tämän tutkimuksen primaari ja sekundaariaineisto on kerätty useassa vaiheessa sekä keskenään rinnakkain. Hirsjärvi ja Hurme (2011) korostavat, että tällaisissa tapauksissa myös aineiston analyysi tapahtuu usein tutkimusaineiston keräämisen yhteydessä. Myös tässä tutkimuksessa, aineiston analysointi tapahtui lomittain sen keräämisen kanssa. Haastatteluaineiston tapauksessa aineiston käsittely aloitettiin niin pian haastattelun jälkeen kuin se oli mahdollista, kuitenkin aina saman päivän aikana. Käsittely aloitettiin haastattelunauhoitteen kuuntelemisella ja puhtaaksi kirjoittamisella, eli litteroinnilla. Rowleyn (2012) mukaan litterointi toimii itsessään alustavana aineiston analyysinä. Tutkijan kuunnellessa äänitettä aineisto alkaa tulla tutuksi ja erilaisia ydinpointteja saattaa jo alkaa hahmottua (Rowley 2012). Sekundaariaineiston tapauksessa aineiston ensimmäiset lukukerrat muodostivat litterointia vastaavan esianalyysin. Aineiston lukemisen yhteydessä kirjoitettiin lyhyitä muistiinpanoja aineistosta nousseista huomioista. Sekundaariaineiston tapauksessa yllä kuvattu esianalyysi oli myös siltä kannalta merkityksellinen, että sen perusteella päätettiin, pystyikö aineisto tuomaan tutkimusongelmaan uutta ja merkityksellistä näkökulmaa ja otettaisiinko sitä mukaan lopulliseen tutkimukseen.

Haastatteluaineiston käsittelyssä käytettävä litteroinnin tarkkuus on riippuvainen tutkimuskysymyksen asettelusta sekä käytetyistä tutkimusmetodeista (Hirsjärvi ym. 2009, 222). Vaikka tutkimukseen valitut haastattelut olivat kaikki asiantuntijahaastatteluja ja tutkimuskysymyksen asettelu korostaa enemmän sisältöä kuin diskurssin tyyliä, päätettiin tutkimuksen litterointi silti toteuttaa sanasanaisesti. Ratkaisuun päädyttiin, sillä haastatteluaineistoa auki kirjoitettaessa sanasanainen litterointi todettiin nopeimmaksi tavaksi saada aineisto valmiiksi jatkokäsittelyyn. Tällöin puhekielisten ilmaisuiden tai täytesanojen poistosta ei tarvinnut tehdä päätöstä äänitettä puhtaaksi kirjoitettaessa. Samalla vältettiin riski sisällöllisesti merkittävän kokonaisuuden pois karsiutumisesta, kun haastatteluaineiston alkuperäinen muoto oli näkyvillä myös litteroinnin jälkeen tehtävässä syvemmässä analyysissa.

Tutkimuksen sisällönanalyysissä sovellettiin induktiivista, eli aineistolähtöistä analyysitapaa. Induktiivisessa analyysissa pääpaino on aineistossa. Tämä tarkoittaa, että analysoitavat kokonaisuudet eivät ole ennalta määrättyjä ja teoria rakennetaan aineisto lähtökohtana. Induktiivisuudella viitataan etenemiseen yksittäisistä havainnoista yleisempiin väitteisiin. (Eskola ja Suoranta 1998, 83) Tuomi ja Sarajärvi (2004, 98) kuitenkin huomauttavat, ettei täysin puhdas induktiivinen päättely ole mahdollista, sillä tutkijan ennakkokäsitykset tutkittavasta ilmiöstä sekä käytetyt käsitteet ja menetelmät ovat tutkijasta riippuvia ja vaikuttavat tuloksiin.

Sisällönanalyysi noudatti Milesin ja Hubermanin (1994) kolmivaiheista prosessia aineiston redusoinnista, eli pelkistämisestä, aineiston klusterointiin, eli luokitteluun ja lopuksi abstrahointiin, eli teoreettisten käsitteiden muodostamiseen ja tulkintaan. Redusoinnissa aineistosta karsitaan pois tutkimusongelman näkökulmasta epäolennaiset kokonaisuudet. Luokitteluvaiheessa aineisto järjestetään siinä ilmenevien pääteemojen mukaisesti. Tämän jälkeen pääteemoille mahdollisesti muodostetaan alateemoja, mikäli sellaisia nousee aineistosta selkeästi esille. Abstrahointivaiheessa tiivistetään aineiston keskeisimmät löydökset ja sovitetaan ne muodostettuun teoreettiseen viitekehykseen. Abstrahointivaiheessa tutkija rakentaa, muodostamiaan käsitteitä käyttäen, kuvauksen tutkimuskohteesta. (Miles ja Huberman 1994)

Erääksi haasteeksi muodostui, ilmiötä relevantisti tutkimuskysymyksen näkökulmasta, kuvailevan aineiston luokittelun valinta. Haastatteluaineistosta oli selkeästi eroteltavissa mallia, opettajaa, dataa ja ympäristöä käsittelevät kokonaisuudet. Tämä luokittelu kuitenkin koettiin päätöksenteon kuvaamista korostavan tutkimuskysymyksen näkökulmasta suboptimaaliseksi, joten siitä luovuttiin. Tutkimusraportin tekoälyn päätöksentekoa kuvaavassa luvussa neljä aineisto on pyritty luokittelemaan siten, että tekoälyn päätöksenteon kuvauksesta saataisiin mahdollisimman rikasta. Lopullisessa luokittelussa näkyy hyvin Tuomen ja Sarajärven (2004, 98) huomio puhtaan induktiivisen analyysin mahdottomuudesta. Tutkimuskysymyksen asettamat reunaehdot aineiston käsittelylle näkyvät tutkimuksessa selvästi. Lisäksi tutkimusraportin kappaleessa 4.6.4 kuvattu, aineistosta nostettu, tekoälyn intuitiivinen luonne muodostaa suoran analogian tutkimuksen teoriaan. Vaikkei aineistoa käsitelty tutkimuksessa teorialähtöisesti, oli tekoälyn intuitiivisen luonteen nouseminen tutkimusaineistosta todennäköisesti ainakin osittain teorian ohjaamaa.

3.5 Luotettavuus

Koskinen ym. (2005) painottaa tutkimuksen luotettavuuden arvioinnin olevan osa tieteellistä tutkimusperinnettä. Tutkimuksen luotettavuutta ei tulisi arvioida pelkästään tutkimuksen lopuksi, vaan arvioinnin tulisi olla kiinteä osa tutkimusprosessia (Eriksson ja Kovalainen, 2008). Tutkimuksen luotettavuuden arvioinnilla pyritään varmistumaan tutkimustulosten todenmukaisuudesta, toistettavuudesta ja yleistettävyydestä. Näitä arvioidaan yleensä reliabiliteetin ja validiteetin kautta. Reliabiliteetilla viitataan tutkimuksen mittaustulosten toistettavuuteen ja tuotetun tiedon kongruenssiin, eli yhdenmukaisuuteen. Validiteetti tarkastelee miten tutkimuksessa tuotettu tieto tai tulkinta kuvaavat ilmiötä, jota tutkimuksessa on tarkoitus kuvata. (Eriksson ja Kovalainen 2008; Hirsjärvi ym. 2009) Vaikka edellä mainitut luotettavuuden arvioinnin kriteerit ovat alun perin kehitetty kvantitatiivisen tutkimuksen luotettavuuden arviointiin, voidaan niitä soveltuvin osin hyödyntää myös kvalitatiivisen tutkimuksen arvioinnissa. (Koskinen ym. 2005) Koska laadullinen tutkimus on luonteeltaan selittävää ja riippuvaisempaa tutkijan valitsemasta näkökulmasta, on tutkimuksen validiteetti reliabiliteettiin verrattuna tällöin korostuneemmassa roolissa (Saaranen-Kauppinen ja Puusniekka 2006). Yin (2009) jakaa validiteetin ja reliabiliteetin neljään kokonaisuuteen: konstruktiovaliditeettiin, sisäiseen validiteettiin, ulkoiseen validiteettiin ja reliabiliteettiin.

Konstruktiovaliditeetti tarkastelee tutkimusmetodin soveltuvuutta valitun tutkimusongelman ratkaisuun (Yin 2009). Tähän tutkimukseen valittu tutkimusongelma on luonteeltaan laadullinen ja korostaa sekä kartoittavia, että kuvailevia piirteitä. Jotta valitut tutkimusmetodit tukisivat tätä tavoitetta tulisi niiden luoda mahdollisimman monipuolinen ja kuvaava aineisto ja sen analyysi. Primaariaineisto kerättiin haastattelemalla useampaa aihealueen asiantuntijaa, joista jokainen toi tutkittavaan ilmiöön omaa asiantuntijuuden aluettaan koskevan yksilöllisen panoksen. Myös tutkimuksen sekundaariaineisto kerättiin useammasta lähteestä ja tiedon paikkaansa pitävyys pyrittiin varmentamaan vertailemalla uutislähteiden sisältöä keskenään. Erikssonin ja Kovalaisen (2008) mukaan kvalitatiivista tutkimusta voidaan usein pitää rikkaampana, luotettavampana ja vakuuttavampana, mikäli se perustuu useampaan empiiriseen datankeruu-menetelmään. Yllä kuvaillun monimenetelmällisyyden eli metodisen triangulaation lisäksi tutkimuksen konstruktiovaliditeettia pyrittiin parantamaan lisäämällä tutkimusprosessin läpinäkyvyyttä. Aineiston keruu,

tutkimusprosessin eteneminen ja aineiston analysointi on kuvattu tutkimusraportin kolmannessa luvussa selvästi ja totuudenmukaisesti. Näin metodisten valintojen perusteet ja niiden mahdollinen sopivuus suhteessa käsiteltävään tutkimusongelmaan on asetettu alttiiksi kriittiselle tarkastelulle.

Tutkimuksen sisäinen validiteetti tarkastelee tutkimuksessa tehtyjen tulkintojen sisäistä loogisuutta ja ristiriidattomuutta (Yin 2009, 42 – 43). Hirsjärven ym. (2009) mukaan erityisesti laadullisessa tutkimuksessa tutkijan kyky irtautua omista lähtökohdistaan ja arvomaailmastaan muodostaa haasteen tutkimuksen sisäiselle validiteetille. Hirsjärvi ym. (2009) korostaakin, että laadullisessa tutkimuksessa on usein lähes mahdotonta saavuttaa objektiivista otetta tutkimukseen. Laine (2018, 28) näkee, että tutkimusaineiston rikas dialogi ehkäisee tutkijan esiymmärryksen ja subjektiivisten näkemysten liiallista vaikutusta tutkimusongelmaa koskevien tulkintojen tekemisessä. Tutkimuksen aineistoa on käsitelty rikkaalla tavalla ja primaariaineiston sana-sanallisella-litteroinnilla ja lainaamisella on vähennetty subjektiivisen tulkinnan mahdollisuutta. Keskenään ristiriitaista tietoa ei ole jätetty tutkimuksen ulkopuolelle. Myös aluksi tutkimuksen ulkopuolelle tarkoitettuja kokonaisuuksia on päätetty ottaa mukaan tarkasteluun niiden esiinnyttyä vahvasti primaariaineistossa. Eräs esimerkki tällaisesta kokonaisuudesta on kappaleessa 5.8 käsitelty eettisten huomioiden kokonaisuus, joka ei kuulunut ennen aineiston keruuta tehtyyn tutkimussuunnitelman versioon.

Aineiston rikkaan ja alkuperäisen muodon säilyttävän käsittelyn lisäksi tutkimuksen sisäistä validiteettia on pyritty parantamaan aineistollisella triangulaatiolla. Monimenetelmällisen aineistonkeräämisen voidaan katsoa rikastavan tutkimuksessa käytettävää aineistoa tuomalla eri näkökulmia ja näin pienentämällä puhtaan subjektiivisen tulkinnan riskiä. Tässä tutkimuksessa sekundaariaineistona toimivien reaalimaailman tapahtumien uskotaan validoivan tutkimuksen primaariaineiston perusteella tehtyjä tulkintoja tekoälyn päätöksenteosta.

Tutkimuksen ulkoinen validiteetti arvioi tutkimustulosten yleistettävyyttä (Yin 2009). Kuten Silverman (2010, 11) kuvaili, on kvalitatiivisen tutkimuksen tavoitteena ymmärtää ilmiötä, eikä tarkastella esimerkiksi sen numeerisesti mitattavaa vaikuttavuutta. Tällöin on mahdollista, että tutkimusaineisto nostaa ilmiöstä esille kokonaisuuksia, joita ei ilmenisi toisessa tutkimusaineistossa. Kvalitatiivisen tutkimuksen pyrkimyksenä onkin analyyttinen yleistäminen, joka tavoittelee teorioiden yleistämistä ja laajentamista

(Saarela-Kinnunen ja Eskola, 2010). Tämän tutkimuksen tarkoitus on yleistää ja laajentaa jo tiedeyhteisön validoimaa päätöksenteko kirjallisuutta tarkastelemalla kokonaisuuksia, joihin teoriaa ei aiemmin ole sovellettu.

Reliabiliteetilla viitataan tutkimuksen tulosten ristiriidattomaan toistettavuuteen eri tutkijoiden toimesta ja eri menetelmiä käyttämällä (Hirsjärvi ym. 2009; Yin 2009). Tutkimusalueen kattavalla tarkastelulla ja haastateltavien valinnalla pyrittiin parantamaan tämän tutkimuksen reliabiliteettia. Haastatteluun valituilla henkilöillä oli asiantuntijuuteen edellytettyä osaamista tutkimuskohteena olevasta ilmiöstä sekä useiden vuosien kokemus työskentelystä ilmiöön liittyvällä sovellusalueella. Primaari- ja sekundaariaineiston käymä dialogi taas validoi tutkijan primaariaineistosta tekemiä huomioita. Tässä tutkimuksessa käsitellyt teemat ovat jatkuvassa muutoksessa. Tekoälyyn liittyvä sekä menetelmällinen että laskentakapasiteettiin vaikuttava kehitys muokkaa tutkimuskohteena olevaa ilmiötä. Tämä muodostaa kriittisen seikan tutkimuksen aineiston ja johtopäätösten toistettavuutta arvioitaessa. Tämän tutkimuksen löydökset lisäävät kuitenkin tämän hetken ymmärrystä tutkimuskohteesta.

4 TEKOÄLY PÄÄTÖKSENTEKIJÄNÄ

Tämä luku syventyy tutkimusraportin tekoälyä käsittelevään aineistoon. Luvussa avataan tekoälyn, koneoppimisen ja luonnollisen kielen käsittelyn termit sekä havainnollistetaan niitä tutkimusaineiston kautta. Luvun tehtävä on vastata tutkimuksen toiseen alatavoitteeseen: kuvata tekoälyn päätöksentekoa. Luku sisältää päätelmiä, jotka on tehty tekoälyjärjestelmissä käytettyjen mallien ja menetelmien pohjalta. Kaikki tekoälyä ja koneoppimista koskeva tekninen ja matemaattinen kirjallisuus on esitetty tutkimusraportin liitteissä.

4.1 Tekoälyn käsitteellinen ymmärtäminen

> *”Toi human level -termi siinä usein tulee vastaan. Tekoäly pystyy ratkaisemaan ongelmia, joita aiemmin vain ihminen on pystynyt. Esimerkiksi toi meidän tunnistusdemo. Pikkulapsikin osaa katsoa kuvaa ja sanoa, että poika, tyttö, mummo, pappa, mut koneelle se on ollu vaikeeta. Siinä on ollu niin paljon semmosta vähäpätöstä dataa niinkun yhen pikselin numeroarvo, niin siit ei voi päätellä mitään, et se täytyy nähdä se koko kuva.” (Huttunen 2018)*

Nilsson (2010, 13) määrittelee tekoälyn pitävän sisällään kaikki aktiviteetit, joiden pyrkimyksenä on koneiden älykkääksi tekeminen. Älyllä Nilsson (2010, 13) viittaa toimintoihin, jotka mahdollistavat entiteetin toiminnan vuorovaikutuksessa ympäristönsä kanssa. Myös Russel ja Norvig (1995, 1) korostavat ympäristövuorovaikutusta määrittelemällä tekoälyn tutkimusalueeksi, joka keskittyy ympäristöä havainnoiviin ja siinä toimiviin kokonaisuuksiin. Ympäristön havainnointi pitää kuitenkin sisällään niin monia funktioita, että niiden listaaminen ja erottelu on vaikeaa, eikä välttämättä tarkoituksenmukaista (Nilsson 2010, 13). Koska käsitteiden, kone ja älykkyys, määritteleminen on hankalaa, lähestytään tekoälyn määrittelyä usein myös suoraan sen sovellusalueiden kautta. Tällaisia ovat muun muassa koneoppiminen, automatisoitu päättely sekä tekstin ja puheen tunnistus. (Oxford, 2017) Tekoälyn sovellusalueet, kuten tekstin, puheen tai hahmojen tunnistus ovat kokonaisuuksia, joilla ihmisen on perinteisesti oletettu toimivan konetta tehokkaammin. Tekoälyn käsite ei kuitenkaan suoraan viittaa tiettyihin teknologioihin tai työkaluihin, vaan käsittelee yleisesti toimintoja, jotka ovat perinteisesti olleet koneelle äärimmäisen haastavia.

Tutkimuksen näkökulmasta tekoälyn eriytyminen omaksi alueekseen on edellyttänyt askeleita useilla eri tieteenaloilla. Logiikka, tilastotieteet, psykologia, neurotieteet ja insinööritieteet ovat kaikki vaikuttaneet merkittävästi nykyisin tekoälytutkimuksena tunnettuun alueeseen. Näiden tieteiden näkökulmasta merkittävimmät edistysaskeleet on esitetty liitteessä 1.

Miellämme älykkyyden mielen ominaisuudeksi, johon liittyy tietoisuus ja kyky opittujen asioiden laajaan soveltamiseen. Ihmisten harjoittama päättelyprosessi kuitenkin eroaa huomattavasti tekoälyjärjestelmien datapisteille kohdistettuihin operaatioihin perustuvasta päättelystä. Varsinaisen päättelyn tasolla tekoälyä voi kuvata ennemminkin tilastotieteeksi, jonka sovellusalue on joko jonkin ihmiselle luontaisen ja koneelle haastavan toiminnan tuottaminen, tai valtavia datamääriä edellyttävä ongelman ratkaisu, jota ei voida lähestyä perinteisten tilastollisten menetelmien avulla. Rajanveto tilastotieteen ja tekoälyn välillä onkin käsitteiden menetelmällisen yhtäläisyyden vuoksi lähes mahdotonta. Sekä tilastotieteen että tekoälyn tapauksessa yhden järjestelmän sovellusalue on kuitenkin aina äärimmäisen rajattu. Esimerkiksi kissojen ja koirien kuvia tunnistava tekoälyjärjestelmä ei kykenisi erottamaan, että sille näytetty norsu ei ole kissa eikä koira, saatikka sitten tunnistamaan näytetyn kuvan esittävän norsua. Kaikki tekoälymenetelmät perustuvat datalle tehtäviin manipulaatioihin. Liitteissä 3–5 on esitelty nykymuotoisten tekoälysovellusten kannalta osa keskeisistä menetelmistä.

4.2 Koneoppiminen

Koneoppiminen on tiiviisti tekoälyn yhteydessä kulkeva käsite. Siinä missä tekoälyn käsite keskittyy enemmän teknologialla tuotettavaan lopputulokseen, kuvaa koneoppiminen tapaa, jolla teknologia käsittelee yksittäistä datapistettä ja oppii tuottamaan tekoälyksi mahdollisesti määriteltävän lopputuloksen. Murphy (2014, 1) määrittelee koneoppimisen keinoksi löytää datasta automaattisesti säännönmukaisuuksia. Koneoppiminen on myös ensisijainen työkalu tulevaisuuden ennustamisessa ja päätöksenteossa epävarmuuden vallitessa (Murphy 2014). Louridas ja Ebert (2016, 110) linkittävät koneoppimisen myös liiketoimintaan ja mainitsevat sen näyttelevän tärkeää roolia toimialat leikkaavassa digitalisaatiossa. He uskovat koneoppimistyökalujen hallinnan olevan koko ajan tärkeämpi yrityksen kilpailukykyä määrittävä tekijä. Yleinen ajatus koneoppimisessa on, että algoritmin ohjelmoinnin sijasta kone opettelee

suorittamaan tehtävän esimerkkitapauksia tarkastelemalla, ja sen jälkeen soveltaa oppimaansa dataan, jota sille ei ole aikaisemmin näytetty (Louridas ja Ebert, 2016).

Koneen opettaminen voi olla joko ohjattua tai ohjaamatonta. Ohjatussa oppimisessa opettajana toimiva entiteetti kertoo koneelle oikean vasteen tiettyyn syötteeseen. Opettaja voi tässä tapauksessa olla ihminen tai toinen automatisoitu järjestelmä. Ohjatun oppimisen tapauksessa esimerkiksi kissojen ja koirien kuvia tunnistavalle järjestelmälle tulee kuvan näyttämisen jälkeen kertoa, onko kyseessä kissa vai koira. Kun tätä toistetaan tarpeeksi monta kertaa erilaisilla syötteillä, oppii järjestelmä lopulta omatoimisesti erottamaan kissat koirista. Ohjaamattomassa oppimisessa vastaavaa opettajaelementtiä ei ole. Kissojen ja koirien tunnistuksen esimerkissä tämä tarkoittaisi sitä, että järjestelmä jakaisi sille syötetyt kuvat kategorioihin kuvan ominaisuuksien perusteella. Huomioitavaa on, että tässä tapauksessa järjestelmä ei osaa itse nimetä luomiansa kategorioita kissoiksi tai koiriksi. Huomioitavaa on myös, että molempien opetustyylien tapauksessa kategorioiden luomisen perusteeksi saattaisi muodostua myös jokin kuvissa oleva, ihmissilmälle näkymätön, säännönmukaisuus. Tällöin muun muassa edellisessä kissaesimerkissä lopputuloksena oleva jaottelu saattaisi tapahtua jonkin muun tekijän, kuin kuvassa esiintyvän eläimen perusteella. (Murphy 2014; Haykin 1994) Opetustyylien välisiä eroja käsitellään tarkemmin liitteessä 2. Organisaation A asiantuntija kuvailee heidän ohjattua oppimista hyödyntävän tekstiä kategorisoivan työkalun toimintaa seuraavasti:

> *"We have SaaS-tool where person can make their own model – – put in some sentences, or if you don't have sentences, you can start to go through your documents, and by clicking say that is this and that is that. So you can basically start from nothing, if you want to, or you can put in 100 examples and it will automate things already – – then you give it a document and it will tell all different categories in that document – – it's quite simple tool, but it's doing a lot behind the scenes"*

Eremenko (2018) jakaa opettamisen viiteen eri vaiheeseen.

1) Datan esikäsittely
2) Datan jakaminen opetus- ja testijoukkoihin
3) Opetettavan mallin luominen
4) Mallin opettaminen opetusjoukon datalla

5) Toiminnan varmistaminen testijoukon datalla

Ympäristöstä saatua dataa ei voida usein suoraan käyttää opetettavan järjestelmän opetusdatana. Datan keräysprosessit eivät ole useinkaan tarkkaan valvottuja, jolloin opetukseen käytettävä data saattaa sisältää virheellisiä tai puuttuvia arvoja. Pyle (1999, 9) korostaa lisäksi datan esikäsittelyn tutustuttavan mallin luojaa dataan ja edesauttavan onnistuneen mallin luomista. Organisaation A asiantuntijan kuvauksessa mallin opettaja nimeää kategorian, johon hänen valitsemansa tekstinpätkä kuuluu ja samalla varmistaa, että malliin syötetyt saman kategorian tekstit todella kuuluvat tähän kategoriaan.

Mikäli järjestelmän käyttötarkoituksen sallimissa puitteissa on mahdollista, esikäsitelty data jaetaan opetus ja testijoukkoihin. Opetusjoukon dataa käytetään valitun mallin opettamiseen ja testijoukon mallin suorituskyvyn testaamiseen. Eremenkon (2018) mukaan yleensä noin 80% datasta tulisi käyttää mallin opettamiseen ja 20% testaamiseen. Luku saattaa kuitenkin vaihdella käsiteltävän ongelman ja datan luonteen mukaan. Mallin luomisen jälkeen sitä opetetaan opetusjoukon datalla. Mallin ominaisuudet määrittävät mitä tekijöitä se datasta nostaa ja minkä datan ominaisuuden suhteen malli oppii. Huttunen (2018) huomauttaa, että ennen mallin käyttöön ottoa emme edes aina pysty arvioimaan, minkä muuttujien suhteen järjestelmä tekee johtopäätöksensä. Voimme vain tarkkailla mallin suoriutumista, kun testaamme sitä viimeisessä vaiheessa testijoukon datalla.

Eremenkon (2018) esittämien opettamisen vaiheiden taustalla on oletus, että mallin luoja on samalla mallin käyttäjä, jolloin hänellä on pääsy sekä sovellusalueen dataan, että mallin luomiseen tarvittavaan osaamiseen. Todellisuudessa tekoälyjärjestelmät toimivat usein osana alihankittua teknologiainfrastruktuuria. Esimerkiksi organisaatio A myy työkaluaan kolmansien osapuolten käyttöön. Tällöin mallin luomisen ja datan käsittelyn yhteys rikkoutuu. Eremenkon (2018) kuvailema vaiheiden välinen järjestys ei siis reaalimaailmassa välttämättä toteudu. Esimerkiksi organisaation A asiantuntijan kuvaus osoittaa, että mallin toimintaa ei aina voida testata ennen mallin käyttöönottoa. Asiantuntijan kuvauksessa testaus tapahtuu työkalun käyttäjän validoidessa järjestelmän tietylle dokumentille antamat kategoriat. Näin testausdatan määrä nousee järjestelmän käytön myötä ja mallin tarkkuus paranee, kun mahdollisia virheitä korjataan niiden noustua esiin. Datan irtaantuminen mallin luomisesta tarkoittaakin usein myös opetuksen

irtaantumista mallin luomisesta, eikä mallin tehneellä toimijalla aina ole kontrollia siihen, miten mallia opetetaan.

4.3 Malli, data ja ympäristö

Tekoälyjärjestelmän päättelystä vastaa usein jokin koneoppiva tilastollinen malli. Eräitä tällaisia malleja on esitetty tutkimusraportin liitteissä 3, 4 ja 5. Käsiteltävä ongelma ja saatavissa oleva data määrittävät millä menetelmillä ongelmaa lähdetään ratkaisemaan. Huttunen (2018) kuvaa mallien valintaa seuraavasti:

> *”Neuroverkothan nyt on ne kuumimmat mistä eniten puhutaan, mutta se riippuu ihan ongelmasta mitä kannattaa käyttää. Et mää opetan omilla kursseilla tulevia datainsinöörejä ajattelemaan avoimin mielin, eikä lukkiutuvan neuroverkkoihin. Monet tuntuu tekevän sitä, et ne näkee vaan et nää neuroverkot on kuuma juttu ja ne on parempia. Tokihan niillä on tehty hienoja juttuja, mutta se ei tarkota sitä, et se ois just sun ongelmaan se paras ratkaisu. Et täytyis pyrkiä käyttämään työkaluja, joissa on helppo kokeilla eri menetelmiä ja iteroida koko ajan.”*

Valittavan mallin lisäksi tekoälyjärjestelmän onnistuminen sille suunnitellun tehtävän suorittamisessa riippuu vahvasti siitä, kuinka hyvin järjestelmän opettamisessa käytetty data tukee tehtävän suorittamista ja kuinka hyvin se kuvaa ympäristöä, jossa tehtävä tulee suorittaa. Organisaation A asiantuntija kuvailee käytettävän opetusdatan ja ympäristön suhdetta kissojen luokittelu -esimerkin kautta seuraavasti:

> *” If you want to classify cats and you take thousands of cat images and train the model with them and in the end have perfect cat classifier and then the user starts to load cat images from the internet, the distribution is quite different. The cats what you had in your data set, doesn’t represent the cats that come from other sets, so this mismatch of distribution is quite difficult. Cat classifier trained with certain distribution is different than classifier trained with other. It doesn’t make sense to use domestic cat classifier to big cats. This is something that we need to explain also to our customers. You need always be inputting text what represents what engineering text is. You can’t train everything, let say with movie scripts and then try to classify engineering text. All the vocabulary is different, user grammar is different,*

> *and all these things differ. Basically, it is so, that if you train on domain A it will work on Domain A."*

Kuten kappaleessa 4.2 todettiin, tekoälyn datasta löytämät säännönmukaisuudet eivät välttämättä vastaa niitä säännönmukaisuuksia, jotka ihminen löytäisi samasta aineistosta ja joiden löytäminen olisi tehtävän kannalta merkityksellistä. Yllä olevassa lainauksessa tämä tulee ilmi asiantuntijan verratessa kotikissojen ja isojen kissojen osajoukoilla opetettuja malleja sekä kuvauksessa insinööritekstille tyypillisten piirteiden merkityksestä. Objektien tunnistus on ihmiselle usein intuitiivista ja kuuluu siten kappaleessa 2.4 kuvaillun systeemi 1:n alle. Ihmiset esimerkiksi tunnistavat kissan piirteet hyvin laajasti erilaisista osajoukoista, vaikkeivat olisi aiemmin nähneet kyseisiä kissoja. Samaten pystymme tunnistamaan, vaikkakin emme välttämättä intuitiivisesti, samaa aihetta käsitteleviä tekstejä riippumatta tekstin tyylilajista. Pystymme esimerkiksi erottelemaan tekstistä teknologiaa käsittelevät kokonaisuudet, riippumatta siitä luemmeko scifi-fantasiaa vaiko tieteellistä tutkimusraporttia.

Jotta kaikki käsiteltävälle sovellusalueelle ja ratkaistavalle ongelmalle relevantit osajoukot saataisiin opetusdatassa kattavasti edustetuiksi, on datan määrän perinteisesti tullut olla huomattava. Pienemmällä datamäärällä toimivien mallien kehittäminen on kuitenkin yksi keskeisistä teknologian kehittämisen tavoitteista. Organisaation A asiantuntija kuvailee heidän ratkaisuaan opetusdatan määrän hillitsemiseksi seuraavasti:

> *"And this is where we're using transfer learning at the moment. So, you need one hundred examples which makes it a lot easier. Normally you'd need some 1000 examples and it's really hard to make a tool that is user friendly, if customers have to spend many hours to teach your tool."*

Siirto-oppiminen, johon Organisaation A asiantuntija viittaa, tarkoittaa koneoppimisen aluetta, jolla yhden ongelman ratkaisemisesta saatua tietoa hyödynnetään toisen, vastaavanlaisen ongelman ratkaisemiseen näin pienentäen vaadittavan opetusdatan määrää (Torrey ja Shavlik, 2009, 1-3). Esimerkiksi kielen rakenteeseen liittyvän ongelman ratkaisemisesta saatavaa tietoa voidaan hyödyntää luokittelu ongelmassa, jossa sisällön ymmärtäminen nousee merkityksellisemmäksi.

Sovellusalueesta vaadittavan datan määrään voidaan menetelmällisten tekijöiden lisäksi vaikuttaa keinotekoisesti generoimalla opetusdataa. Eräs esimerkki tästä on tutkimusraportin viidennessä luvussa paremmin esiteltävät Facebookin neuvottelemaan

opetetut chat-botit, joiden kouluttamista varten kirjoitettiin yli 5000 kuvitteellista dialogia. Opetusdatan generoiminen saattaa kuitenkin joissain tilanteissa johtaa Organisaation A asiantuntijan kissojen luokitteluesimerkillä havainnollistamaan tilanteeseen, jossa generoitu data ei enää kuvaa realistisesti sovellusaluetta.

4.4 Luonnollisen kielen käsittely

Luonnollisen kielen käsittely *(natural language processing)* viittaa tekoälyn osa-alueeseen, joka tarkastelee, kuinka koneita voidaan käyttää ymmärtämään luonnollista kieltä (Chowdhury 2003, 51). Luonnollisen kielen käsite erottaa ihmisten puhumat kielet koneen ymmärtämistä kielistä ja sisältää sekä puhutun että kirjoitetun kielen (Nilsson 2010, 141). Koneellinen kielen käsittely lähtee kielen ja merkitysten synnyttämisen rakenteesta. Kieltä voidaan analysoida useilla eri tasoilla, jotka voidaan jakaa keskinäisiin hierarkioihin. Alimmilla tasoilla ovat äänteet ja sanan osat. Ylemmät tasot liittyvät kielen jaksotukseen ja lauseiden rakenteeseen. (Nilsson 2010, 141) Nilsson (2010, 142) listaa viisi erilaista kielen tasoa:

1) Fonologinen taso
2) Morfologinen taso
3) Syntaksinen taso
4) Semanttinen taso
5) Pragmaattinen taso

Fono- ja morfologiset tasot käsittelevät sanan muodostusta. Fonologia jakaa sanat kielen äänteisiin ja tutkii, miten merkityksiä muodostetaan äänne-erojen kautta. Tämä taso on merkittävä puheentunnistuksessa, jossa puhuttu kieli muutetaan ennen jatkoprosessointia tekstimuotoon (Nilsson 2010, 254). Morfologia tarkastelee sanojen muodostumista pienemmistä kokonaisuuksista. Esimerkiksi sana talossa muodostuu perusmuodosta talo ja -ssa päätteestä.

Syntaksinen taso käsittelee kielen sääntöjä, se määrittelee reunaehdot, jotka sanojen muodostaman jonon tulee täyttää, jotta sitä voidaan pitää kieliopillisesti oikeana. Yhdessä sanojen määritelmien kanssa syntaksi muodostaa fono- ja morfologiaa ylemmän tason lauseen merkityksen ymmärtämisessä. Pelkkä kieliopillinen oikeellisuus ei kuitenkaan puhtaasti riitä määrittelemään lauseen merkitystä. Lause voi olla kieliopillisesti oikein mutta järjetön.

Semanttinen taso tarkastelee kieltä merkitysten kautta. Ilmaisujen välisten suhteiden tarkastelu muodostaa semanttisen tason ytimen. Kielen ylin, pragmaattinen taso taas tulkitsee merkitystä laajemman kontekstin kautta. Esimerkiksi ilmaisun, ”räjäyttää pankki”, merkitys vaihtuu tarkasteltavan kontekstin myötä. (Nilsson 2010, 142-147)

Manningin (1999, 8) mukaan kaksi keskeisintä kysymystä, joihin kielitieteilijät pyrkivät vastaamaan ovat: Millaisia asioita ihmiset sanovat ja mitä nämä asiat kertovat ympäröivästä maailmasta. Tekoälyjärjestelmien näkökulmasta ensimmäisen kysymyksen voidaan nähdä tarkastelevan sitä, millaisia sanoja ihmiset käyttävät ja jälkimmäisen, mikä on sanoista muodostettujen lauseiden syvempi merkitys. Koska ihmiset aina venyttävät kielen sääntöjä ja rajoja, on algoritminen lähestyminen kielen merkityksiin haastavaa (Manning 1999, 3). Tämän vuoksi luonnollisen kielen käsittelyn ongelmat käsittelevät usein ensimmäistä kysymystä. Tämä kysymys voidaan muotoilla käsiteltävän sovellusalueen mukaan: Millaisia asioita ihmiset sanovat, kun käsittelevät sovellusaluetta x. Käytettyjen sanojen ja sovellusalueiden suhteita voidaan sitten opettaa automatisoiduille järjestelmille ja näin saavuttaa ymmärrystä käsiteltävistä alueista.

Kieleen liittyvä tulkinnanvaraisuus nähdään keskeisenä ongelmana teknologiaa kehitettäessä. Organisaation A asiantuntija kommentoi kielen ymmärtämisen haasteita seuraavasti:

> *” That is actually one of the problems with this tech, that it’s meant to be really clear and understandable, but the problem is that it actually is not. You have different cultures, and different words and all that stuff. There are certain syntaxes on technical document what we are looking, for example there are models which basically tell how you are supposed to write a sentence – – and yet people complain all the time that they get misunderstandings. So, it’s surprisingly bad just because it’s language. It will always have these issues.”*

Erityisesti kielen syntaksinen taso on ollut merkittävässä roolissa luonnollisen kielen käsittelyn aikaisemmassa tutkimuksessa ja muodostaa edelleen tärkeän osan luonnollisen kielen käsittelyn sovelluksista (Nilsson 2010, 142). Automaattinen kielen kääntäminen ja luonnollisen kielen käyttäminen koneen käyttöliittymän ohjaamiseen olivat ensimmäisiä luonnollisen kielen käsittelyn tutkimusta vauhdittavia tavoitteita. Niiden rinnalle on

noussut kvalitatiivista, tekstimuotoista tietoa ymmärtävien ja sitä tiivistävien järjestelmien kehittäminen. (Nilsson 2010)

4.5 Tilastollisten mallien suoriutuminen ihmiseen verrattuna

Tekoälyjärjestelmien taustalla olevien tilastollisten järjestelmien suoriutumista ihmiseen verrattuna on käsitelty aiemmissa tutkimuksissa. Tässä kappaleessa esitellään kirjallisuutta, jossa tekoälyjärjestelmien kannalta keskeisiä menetelmiä – neuroverkkoja, lähimmän naapurin menetelmää ja päätöspuita – vertaillaan inhimillisen päätöksenteon heuristisiin malleihin. Yllä olevat menetelmät on esitelty tutkimusraportin liitteissä 3–5.

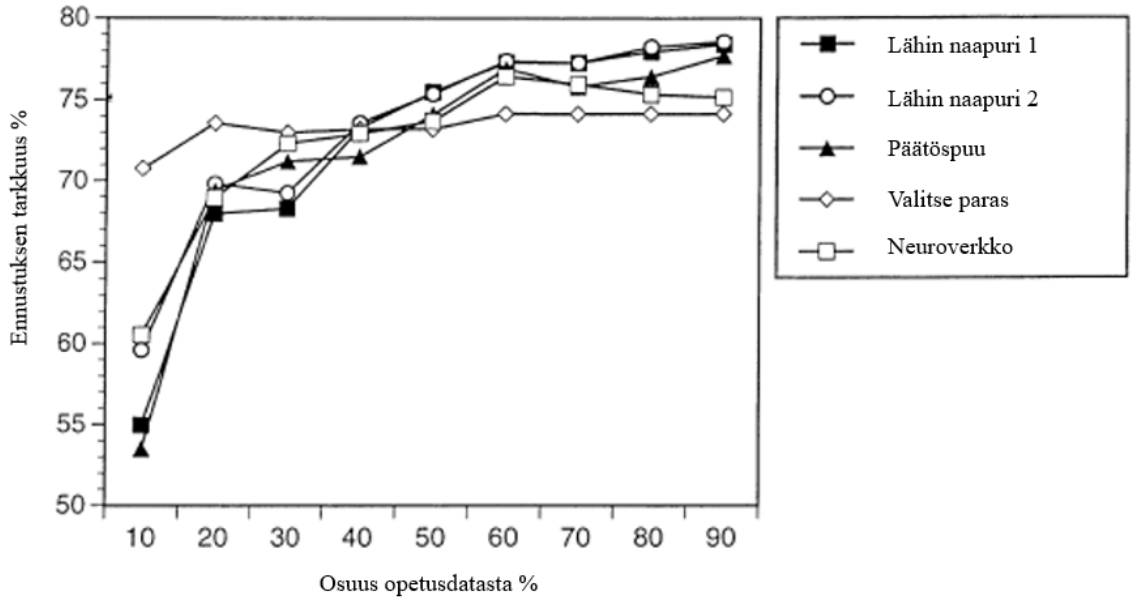
Tutkimusraportin kappaleessa 2.5 viitattiin lyhyesti Meehlin (1954) julkaisuun, joka kävi läpi 20 eri tutkimusta, joissa kliinisen psykologin kykyä ennustaa yksilön käyttäytymistä verrattiin tilastollisen malliin vastaavaan kykyyn. Meehlin (1954) esittelemän aineiston mukaan tilastolliset mallit ennustivat käyttäytymistä ylivertaisesti ammattilaisen arvioon verrattuna. Myös päätöksenteon vinoumia käsittelevä kirjallisuus lähestyy päätöksentekoa usein ihmisen rajoittuneiden tilastollisten kykyjen kautta ja antaa olettaa, että turvaudumme yksinkertaistuksiin ja pienempiin datamääriin, vain koska parempien menetelmien edellyttämää laskentatehoa ei päätöshetkellä ole saatavissa, tai sen käyttö edellyttää tilanteen merkittävyyteen nähden liian suuria kognitiivisia ponnistuksia.

Gigerenzer ja Brighton (2009) esittävät, että vastoin yleistä oletusta, käytettävän datan tai laskentatehon tai -ajan lisääminen eivät kaikissa tilanteissa paranna menetelmän luotettavuutta. Heidän mukaansa heuristiikat sekä nopeuttavat päätelmien tekemistä, että saattavat lisätä niiden tarkkuutta. Heidän mukaansa heuristiikat nopeuttavat päätöksentekoa pääsääntöisesti kahdella tavalla: jättämällä huomioimatta joko päätöstilanteeseen vaikuttavia muuttujia tai muuttujien välisiä eroja niiden painoarvoissa.

Gigerenzer (2008, 21) viittaa muun muassa Czerlinskin ym. (1999) tutkimukseen, jossa verrattiin kolmen eri heuristiikan suoriutumista regressioanalyysiin 20:ssä erilaisessa ennustamistehtävässä. Czerlinskin ym. (1999) tutkimuksessa eri ennustamistehtävien havainnot vaihtelivat 11:sta havainnosta 395:een ja kolmesta ennakoivasta muuttujasta 18:ta. Regressioanalyysiin verrattavat heuristiikat olivat valitse-paras-heuristiikka (*Take the best),* joka esiteltiin tutkimusraportin kappaleessa 2.6, Minimalisti-heuristiikka *(Minimalist)* ja vakio-painoarvo-heuristiikka *(Tallying).* Minimalisti-heuristiikka on yhteneväinen valitse-paras-heuristiikan kanssa sillä erotuksella, että vihjeiden

validiteettijärjestyksen sijasta vihjeet ovat sattumanvaraisessa järjestyksessä. Vakio-painoarvo-heuristiikka viittaa menetelmään, jossa kaikkia valintakriteeriä ennustavia muuttujia arvioidaan samalla painoarvolla jättäen mahdolliset erot painoarvoissa huomioimatta. Czerlinskin ym. (1999) tutkimuksessa havaittiin, että kaikki heuristiikat onnistuivat ennustamaan, tarkastelun kohteena olevaa muuttujaa, regressioanalyysiä paremmin kaikissa paitsi yhdessä tehtävässä.

Chater ym. (2003) saivat Czerlinskin ym. (1999) kanssa samansuuntaisia tuloksia tutkimuksessaan, jossa he vertailivat valitse-paras-heuristiikkaa kahta erilaista lähimmän naapurin menetelmää, päätöspuuta sekä kolmikerroksista, eteenpäin syöttävää neuroverkkoa. Chaterin ym. (2003) tutkimuksessa tehtävänä oli yhdeksän muuttujan avulla valita kahdesta esitetystä kaupungista suurin. Chaterin ym (2003) tutkimuksessa tilastollisia malleja opetettiin 6806 erilaisella, kahdesta kaupungista muodostetulla, parilla. Opetusdataa oli siis huomattavasti enemmän kuin missään Czerlinskin ym. (1999) tutkimuksen ennustamistehtävässä. Kuviossa 5 esitetään mallien suoriutuminen Chaterin ym. (2003) tutkimuksessa. X-akseli kuvaa käytetyn opetusdatan määrää ja Y-akseli onnistuneiden suoritusten osuutta prosentteina.



Kuvio 5: Mallien suoriutuminen (Chater ym. 2003, 77)

Kuviosta 5 nähdään, että valitse-paras-heuristiikka suoriutuu parhaiten, kun opetukseen käytettävä datamäärä on rajallinen. Muut menetelmät kuitenkin ohittavat valitse-paras-heuristiikan ennustustarkkuudessa, kun opetusdatan määrää lisätään.

Taulukkoon 4 on listattu heuristiikkoja ja tilastollisia menetelmiä vertailevat tutkimukset. Tässä kappaleessa jo tarkemmin esiteltyjen tutkimusten lisäksi taulukkoon on nostettu myös Hogartin ja Karelaian (2007) sekä Brightonin (2006) tutkimukset, joiden tulosten voidaan kuitenkin katsoa olevan hyvin linjassa Chaterin ym. (2003) sekä Czerlinskin ym. (1999) kanssa.

Taulukko 4: Heuristiikkoja ja tilastollisia menetelmiä vertailevat tutkimukset

Tutkimus	**Kuvaus**	**Löydökset**
Meehl (1954)	Julkaisu, joka vetää yhteen 20 eri tutkimusta, joissa tilastollista ennakoimista verrataan asiantuntijan tekemään ennustukseen	Tilastolliset mallit suoriutuivat ylivertaisesti asiantuntijan päätökseen verrattuna
Czerlinski, Gigerenzer, Goldstein (1999)	Verrattiin kolmen eri heuristiikan suoriutumista lineaariseen regressioon 20:ssä erilaisessa ennustustehtävässä	Heuristiikat suoriutuivat ylivertaisesti lineaariseen regressioon verrattuna
Chater, Oaksford, Nakisa, Redington (2003)	Verrattiin valitse paras -heuristiikkaa, kahta lähimmän naapurin menetelmää, neuroverkkoa ja päätöspuita. Tehtävässä, tuli valita kahdesta esitetystä kaupungista suurin yhdeksän esitetyn muuttujan avulla.	Valitse-paras-heuristiikka suoriutui muita malleja paremmin kapealla opetusdatalla, mutta jäi tarkkuudessa jälkeen, mikäli opetusdataa lisättiin.
Hogart ja Karelaia (2007)	Verrattiin valitse paras -heuristiikkaan verrattavissa olevaa yhden muuttujan menetelmää ja vakiopainojen käyttöä regressioanalyysiin neljässä simuloidussa ennustamistehtävässä	Yhden muuttujan menetelmä ennusti parhaiten kahdessa ympäristössä. Regressioanalyysi sekä vakiopainojen käyttö nousivat molemmat parhaaksi yhdessä ympäristössä. Mallien väliset erot kuitenkin hillittyjä.
Brighton (2006)	Verrattiin valitse-paras-heuristiikkaa, kahta erilaista lähimmän naapurin menetelmää, päätöspuuta ja neuroverkkoa kahdeksassa erilaisessa ennustamistehtävässä.	Valitse-paras-heuristiikka suoriutui muita menetelmiä paremmin 50%:ssa ennustustehtävistä. Lopuissa 50%:ssa suoriutui paremmin vain, mikäli mallien opetusdataa kavennettiin.

Kaikissa listatuissa tutkimuksissa heuristiikat: valitse-paras ja vakio-painoarvo, olivat korostuneessa roolissa. Nämä edustavat Gigerenzerin ja Brightonin (2009) luonnehdintaa heuristisen päätöksenteon taipumuksesta yksinkertaistaa tilannetta joko vihjeiden lukumäärän tai vihjeiden välisten painoarvojen suhteen. Nämä molemmat heuristiikkoihin liittyvät ominaisuudet tulevat, taulukoiduissa tutkimuksissa edustetuiksi. Äärimmäisyyksiin yksinkertaistettujen ympäristöjen vuoksi listatuissa tutkimuksissa vertailtuja tilastollisia ja heuristisia malleja ei kuitenkaan voida pitää kokonaisvaltaisina kuvauksina ihmisen suoriutumisesta konetta vastaan päätöstilanteessa. Tutkimukset kuitenkin osoittavat, ettei heuristinen yksinkertaistus aina tarkoita uhrausta päätöksenteon tarkkuudessa. Tämän lisäksi ne ohjaavat tarkastelun suuntaan. Koska tutkimustulokset vaihtelivat ympäristön ja esitetyn ongelman mukaan, relevantti kysymys ei ole: kumpi suoriutuu paremmin päätöstilanteessa, ihminen vai kone, vaan millaisessa ympäristössä toinen suoriutuu paremmin.

Shanteau ja Thomas (2000, 762) viittaavat Czerlinskin ym. (1999) tutkimukseen ja esittävät kaksi tekijää, jotka ovat yhteisiä kaikille ympäristöille, joissa heuristiikat suoriutuivat tilastollisia menetelmiä paremmin.

1) Alhainen havaintojen määrä suhteessa ennakoiviin muuttujiin
2) Ennakoivat muuttujat yhteisvaihtelevat positiivisesti

Vaikka Shanteaun ja Thomaksen (2000, 762) tekemät havainnot on tehty vain Czerlinskin (1999) tutkimuksen pohjalta, ovat ne linjassa myös myöhemmin julkaistujen tilastollisia menetelmiä ja heuristiikkoja vertailevien tutkimusten kanssa.

4.6 Tekoäly päätöksentekijänä

Edellisen kappaleen perusteella nähdään, että korkea havaintojen lukumäärä suhteessa ennakoiviin muuttujiin muodostaa kriittisen reunaehdon tilastollisten mallien ja sitä kautta myös tekoälyjärjestelmien käyttämiselle. Vähäisten havaintojen ympäristöissä tilastollisen mallin rakentaminen ei ole kannattavaa, sillä heuristinen päätöksenteko saavuttaa tarkemman lopputuloksen pienemmällä laskentateholla. Vaikka edellisessä kappaleessa viitatut tutkimukset antavat hyvän lähtökohdan tarkastella tilastollisia menetelmiä päätöksenteon näkökulmasta, ne eivät juuri tarjoa relevanttia tekoälyn päätöksentekoa kuvailevaa tietoa. Tekoälyjärjestelmien soveltaminen tapahtuu usein joko ympäristöissä, jotka ovat kappaleessa 4.5 kuvattuja ympäristöjä kompleksisempia ja

joissa havaintojen lukumäärä on niin suuri, ettei muu kuin tilastollinen käsittely ole mahdollista, tai yksinkertaisemmissa ympäristöissä, joissa tekoälyä käytetään jonkun ihmiselle luontaisen toiminnan kuten tekstin tai puheen tunnistuksen tehostamiseen tai automatisointiin. Erityisesti jälkimmäisessä tapauksessa tekoälyn suorittamat päätökset saattavat lopputuloksen puolesta olla yhteneväisiä ihmisen vastaavissa tilanteissa tekemien päätösten kanssa. Vaikka lopputulos olisi yhteneväinen on päätöksen syntymismekaniikka kuitenkin niin erilainen, ettei tekoälyn suora rinnastaminen ihmispäättelyyn välttämättä ole tarkoituksenmukaista ja saattaa tietyissä tilanteissa johtaa kriittisiin väärinymmärryksiin.

> *”Oon vetäny muutamia tekoälykoulutuksia, niin se ydinsanoma on se, et älkää verratko ihmisaivojen toimintaa tekoälyn toimintaan, koska se on lähtökohtasesti niin erilaista” (Lehti 2018)*

Tässä luvussa identifioidaan tutkimuksen primaariaineistosta kolme tekoälyn päätöksentekoa kuvaavaa ominaisuutta. Nämä ovat: kohdennettu soveltaminen, rajoitettu deskriptiivisyys ja riippuvuus kolmansista osapuolista. Viimeinen kappale tiivistää tekoälyn päätöksenteon kuvauksen, yhdistää sen sovellusympäristöön ja vertaa tekoälyn päätöksentekoa inhimilliseen päätöksentekoon. Luku vastaa tutkimuksen kolmanteen alatavoitteeseen, kuvata tekoälyn päätöksentekoa.

4.6.1 Kohdennettu soveltaminen

> *”Ärsyttää kun joskus lehdistössä oli artikkeli, et tekoäly on äärettömän typerä – – Ja kun joo, se on kyllä sitäkin, mutta samaan aikaan myös äärettömän älykäs. Se on vaan niin äärettömän kapealla alueella äärettömän älykäs.” (Lehti 2018)*

Tekoäly ei ennen opettamista erittele lopputuloksen kannalta merkityksellisempää informaatiota. Toisaalta vähemmän merkityksellistä dataa ei suodateta käsittelyn ulkopuolelle. Jokaista datapistettä käsitellään identtisesti ja syötedatan painoarvo muodostuu lopulta aina opetuksen lopputuloksena tavoitteena olevaan päämäärään vertaamalla. Kuten koneoppimista käsittelevässä kappaleessa avattiin, tämä datapisteiden yksityiskohtainen käsittely rajoittaa saman tekoälyjärjestelmän soveltuvuutta erilaisten ongelmien ratkaisuun. Lisäksi tietyn ongelman ratkaisemiseen kerätty data on usein niin spesifiä tietylle sovellusalueelle, ettei sama järjestelmä kykene ratkaisemaan vastaavaa

ongelmaa toisella sovellusalueella. Huttunen (2018) nostaa esiin myös ratkaistavaksi halutun ongelman rajaamisen haasteen.

> *”Ongelman rajaus on usein reaalimaailmassa se hankalin tekijä. – – Ja se, että miten muotoilee sen ongelman. Et onko se yksinkertainen kyllä, ei, viallinen vai ehjä vai ennustetaanko vaikka jonkun komponentin jäljellä olevaa käyttöaikaa, joka ois varmaan paljon relevantimpi. Mutta paljon vaikeempi kerätä se aineisto. Et pitäs niinkun ottaa kuva nyt ja sitten oottaa viis vuotta ja kattoo miten kävi.” (Huttunen)*

Ongelman rajaamista hankaloittaa tarpeeksi spesifin datalta kysyttävän kysymyksen hahmottelun lisäksi ongelman ratkaisemiseen vaadittavan datan kerääminen. Huttunen (2018) jatkaa datan keräämiseen liittyvää pohdintaa seuraavasti:

> *”Datan keräys on suurin kustannustekijä. Fysikaalisen maailman tilanteissa vielä enemmän, kuin esimerkiksi jonkun internetistä kerättävän aineiston tai taloudellisen aineiston kerääminen, kun sitä nyt kerätään muutenkin ja sitä on jo olemassa. Mutta jos halutaan nyt vaikka tutkia komponenttien kestävyyttä, niin se täytyy tehdä siellä ja yksilöidysti siellä. Sitä dataa ei muualla kerätä.”*

Ongelman ratkaisu voi edellyttää datan keräämisen lisäksi myös täysin uuden datan luomista. Eräs tällainen esimerkki on Facebookin tekemä neuvottelevia tekoälybotteja käsittelevä tutkimus. Neuvottelevan tekoälyn tutkimista varten luotiin 5808 ihmisten käymää dialogia kuvitteellisesta tilanteesta, jossa jaettiin kirjoja, hattuja ja palloja kahden neuvottelijan välillä (Lewis ym 2017). Facebookin neuvottelevia tekoälybotteja käsitellään uudelleen tämän tutkimusraportin kappaleessa 5.6. Mikäli datan kerääminen tai relevantin datan luominen on mahdollista ja ongelma saadaan rajattua, näkee Huttunen ratkaistavissa olevien ongelmien joukon kuitenkin laajana:

> *”Sanosin, että mitä tahansa voidaan ratkaista, jos on rahaa. Et ollaan nyt siirtymässä kohti halvempia ja halvempia ratkaisuja ja pystytään ratkaisemaan taloudellisessa mielessä vähäpätöisempiäkin ongelmia.”*

Tiedusteltaessa mahdollisuuksista laajentaa tekoälyjärjestelmän käsittelemien ongelmien joukkoa, Huttunen (2018) mainitsee auto ML algoritmit ja kuvailee niitä seuraavasti:

> *”Lähimmät mihin siinä suunnassa on menty, on tämmöset Auto-ML (automatic machine learning) algoritmit, – – et nekin on vaan yleistyksiä siitä yhteen ongelmaan opettamisesta. Et annetaan kasa ongelmia ja sinne sitten rakenteita, jotka osaa päätellä, et miten tää nyt sitten ratkastas. Mut nääkin on kuitenkin kapeita ja aina kategorisia, et ne pystyy laajentamaan ehkä yhdestä ongelmasta kymmeneen ongelmaan, jotka on rajattuja ja hyvin määriteltyjä, mikä on usein reaalimaailmassa vaikea saavuttaa.”*

Huttusen kuvaamat auto-ML algoritmit jakavat yhteispiirteitä kappaleessa 4.3 lyhyesti sivutun siirto-oppimisen kanssa, jossa yhden ongelman ratkaisemiseksi nähtyä opetustyötä voidaan hyödyntää järjestelmän opettamisessa toisen vastaavanlaisen ongelman ratkaisuun. Siirto-oppimisessa ratkaistavien ongelmien tulee kuitenkin jakaa merkittäviä yhdistäviä piirteitä, jotta menetelmän hyödyntäminen olisi mahdollista. Auto-ML algoritmien tapauksessa Huttunen (2018) korostaa käsiteltävien ongelmien kapeutta, kategorisuutta ja selkeää rajausta.

Tekoälyn edellyttämää tiukkaa rajausta ja selkeää määrittelyä kutsutaan tässä tutkimuksessa kohdennetuksi soveltamiseksi. Tämä reunaehto koskee käsittelyssä olevaa sovellusaluetta, ratkaistavaa ongelmaa ja ongelman ratkaisuun kerättävää dataa. Tekoäly kykenee tehokkaaseen päätöksentekoon vain kapealla, selkeästi rajatulla sovellusalueella, josta kerätty data tukee suoraan ratkaistavaksi valittua, selkeästi määriteltyä tehtävää.

4.6.2 Rajoitettu deskriptiivisyys

> *”Ongelma on se, että deterministisiin malleihin verrattuna, missä sulla on joku kaava, sä pystyt hyvin ennustaan, miten se kaava toimii mutta noi tekoälyt on lähtökohtasesti sellaisia, että ne on ei deterministisiä, että sä pystyt hyvin hassuilla asioilla johtamaan sitä harhaan ja sä et pysty oikein ennustamaan missä kohtaa sä johdat sitä harhaan.” (Lehti 2018)*

Kykymme ymmärtää ja ennakoida tekoälyn päätöksentekoa on rajallinen. Pystymme kuvailemaan ja selittämään itse mallin toimintaa, tarkkailemaan sen datalle tekemiä operaatiota ja täten seuraamaan tekoälyn ”päättelyä” yksittäisten datapisteiden tasolla, kuten esimerkiksi liitteissä 3–5 on esitetty. Tämä ei kuitenkaan auta meitä muodostamaan loogisesti ymmärrettävää selitystä tekoälyn tekemistä päätöksistä. Voimme arvioida

järjestelmän toimintaa tarkkailemalla sen suunniteltuun ongelmaan tuottamaa lopputulosta, mutta emme pysty kyseenalaistamaan sen päättelyketjua.

Selitettävyyden ongelma liittyy päättelyketjun lisäksi sekä yksittäisiin tapahtumiin että tehtäviin. Kaksi tapahtumaa tai tehtävää, jotka näyttäytyvät ihmiselle identtisinä eivät välttämättä näyttäydy niin tekoälylle. Lehti (2018) kuvailee tähän liittyvää problematiikkaa seuraavasti:

> *”Tekoälykin, kun sille annetaan sama input niin siitä tulee aina sama output. Paitsi tämmösissä tapauksissa, joissa on esimerkiks kuva tai videomateriaalia, et jos laitat saman tyypin liikkumaan kameran edestä, niin vaikka ne on tavallaan samat tapahtumat, niin ne ei inputin suhteen silti välttämättä ole. Ja sitten kun sä et pysty näkemään sitä, et täähän on tää, mikä se on.”*

Huomioitavaa on, että vaikka ihmisen ja tekoälyn tulkinta tapahtuman eroavaisuudesta voi olla erilainen ei tämä tarkoita, että se kaikissa tapauksissa on. Tekoälyjärjestelmien kehittämisen pyrkimyksenä onkin usein löytää tarpeeksi yhteneväinen tulkinta ihmisen ja koneen välillä, jotta käsiteltävän ongelman ratkaiseminen olisi mahdollista. Tulkinnan yhteneväisyyttä voidaan kuitenkin tarkastella vain testaamalla järjestelmän toimintaa erilaisissa ympäristöissä. Tällöin emme pysty koskaan täysin varmasti sanomaan, onko tulkintamme todellisuudessa yhteneväinen järjestelmän kanssa, vai näyttäytyikö se vain yhtenäisenä siksi, että tekoälyn tulkinta on tuottanut saman lopputuloksen siinä rajatussa määrässä ympäristöjä, jossa olemme sitä testanneet.

Edellä kuvattua tapahtumiin, tehtäviin ja päättelyketjun läpinäkyvyyteen liittyvää selitettävyyden ja ymmärrettävyyden ongelmaa kutsutaan tässä tutkimuksessa rajoitetuksi deskriptiivisyydeksi. Rajoitettu deskriptiivisyys voidaan nähdä myös eräänä rajoitettua sovellettavuutta selittävänä tekijänä. Koska asiat ja ongelmat, jotka ihminen tulkitsee samanlaisiksi, eivät tekoälyn näkökulmasta välttämättä näyttäydy samanlaisina, minimoidaan teknologiaa kehitettäessä tekijät, jotka tunnistamme erilaisiksi. Nämä tekijät rajataan opetettavan järjestelmän käsittelyn ulkopuolelle ja toivotaan, että jäljelle jäävä kokonaisuus näyttäytyy sekä tekoälyjärjestelmän että ihmisen näkökulmasta tarpeeksi samanlaisena, jotta käsiteltävän ongelman ratkaiseminen on mahdollista. Esimerkiksi koneoppimista käsittelevässä kappaleessa organisaation A asiantuntija kuvaili heidän tuotteensa edellyttämää opetussyötettä seuraavasti:

> *"You need always be inputting text what represents what engineering text is. I can't train everything, let say with movie scripts and then try to classify engineering text. All the vocabulary is different, user grammar is different, and all these things differ."*

Organisaatio A pyrkii rajaamaan opetussyötettään siten, että merkitsevimmäksi erottavaksi tekijäksi eri syötteiden välillä muodostuisi tekstin sisältö ja muut eroavaisuudet kuten sanasto ja lauserakenne minimoitaisiin.

Rajoitettu deskriptiivisyys näkyy teknologian kehittämisessä myös mallien luomisen tasolla. Koska tekoälyn löytämät säännönmukaisuudet ja erottavat tekijät eivät välttämättä vastaa ihmisen samasta aineistosta löytämiä, vaan perustuvat osaltaan arvaukseen, muodostuu myös mallien luominen hyvin kokeilevaksi. Huttunen (2018) kuvaa mallien luomista seuraavasti:

> *"Koko ajan täytyy olla kolme mallia valmistumassa ja tulkita siitä sitten mikä on paras. – – et se teoria on toisarvosta, kun tehdään asioita, joiden pitää toimia. Et sitten vaan kokeillaan kaikkia, vaikka teorian mukaan tän ei pitäskään olla hyvä tämmösissä ongelmissa, mut kaikki käydään läpi ja valitaan se paras, et se ei oikeastaan niin paljoa kiinnosta et miks se toimii, kunhan se toimii ja asiakas maksaa laskun."*

Huttunen (2018) jatkaa mallien testaamisen kuvailua:

> *"Kokeillaan ottamalla hyvin menestyneitä verkkoja, joita jossain kilpailuissa on käytetty ja kokeillaan, et kuis hyvä tää olis. Tai sitten, että ihan tylysti randomilla generoi verkkoja, et satunnaisesti arpoo, et vaikka neljä konvoluutio leijeriä ja vaikka kaks dense leijeriä ja vaikka tommonen epälineaarisuus ja noin monta nodee ja kaikki nää arpoo ja jättää viikonlopuks pyörimään ja kattoo kuinka hyvä tulos tuli ja ottaa niistä sen parhaan."*

Kokeileva kehitystyö asettaa huomattavan paineen testaamiselle. Jotta voisimme varmistua, että tekoälyn löytämä säännönmukaisuus todella vastaa säännönmukaisuutta, jonka toivomme löytävämme tai tuottaa toivotun tuloksen kaikissa tehtävän kannalta relevanteissa ympäristöissä, tulee meidän kiinnittää erityistä huomiota testausjoukkojen

valitsemiseen. Huttunen (2018) kuvaa joukkojen valinnan ja mallin testaamisen merkitystä suraavasti:

> *”Sit täytyy olla vaan kieli keskellä suuta, kun valitsee nää joukot (testausjoukot). Et voi käydä niinkin et näyttää ihan samanlaisia esimerkkejä. Et esimerkiksi, vaikka jossain audiosignaalissa, että näyttää yhden pätkän ja testaa sen toimintaa sitten heti seuraavalla pätkällä, joka on melkein samanlainen. – – Evaluointi on tärkee ja siihen ei ehkä ihan tarpeeks panosteta koulutuksessa tällä hetkellä ja se mitä siitä seuraa niin on, että suunnittelee mallin, testaa sitä virheellisesti, saa tosi hyvät lukemat ja lopulta ampuu itteensä jalkaan kun sanoo asiakkaalle, et täs on teille tää ratkasu ja asennetaan se asiakkaan tuotantoon ja käy ilmi, et se ei toimi ollenkaan, se on nolo tilanne.”*

Jotta Huttusen (2018) kuvaama virheellisen testauksen ansa vältettäisiin, tulisi mallia testatessa altistaa se kaikille käyttöympäristössä mahdollisesti ilmeneville tilanteille. Koska ihmisen ja tekoälyn tulkinta käyttöympäristöstä on erilainen, on meidän kuitenkin hankala tietoisesti valita tilanteita, jotka ovat tekoälyn tulkinnan kannalta merkitseviä. Myös mielikuvituksemme asettaa usein rajoitteen sille kuinka laajan mahdollisten tapausten joukon pystymme kustakin käyttöympäristöstä keksimään.

4.6.3 Riippuvuus kolmansista osapuolista

> *”Tekoälyllä, vaikka siin ei olis ihmistä opettamassa, niin ne tilanteet mille se tekoäly, vaikka itseoppivakin tekoäly tulisi altistumaan. Esim. suomessakin ne, jotka on teknisesti lahjakkaita tai orientoituneita, niin tulee enemmän kanssakäymiseen sen tekoälyn kanssa, jolloin syntyy systeeminen bias sen kautta. Se järjestelmä ei tule näkemään tiettyä osajoukkoa ihmisistä ja sillä kouluttajalla on osansa tässä, että jos kouluttaja ymmärtää biaksia, niin se pystyy estämään niitä ja sit kun se kone oppii itse, niin sen täytyis jotenkin varmistaa, et se pääsee osaksi tarpeeksi isoon joukkoon, ettei sitä biasta synny.”*

Tekoäly ei toimi itsenäisesti, vaan on riippuvainen sitä opettavista ja sen tulosta tulkitsevista tahoista. Kuten edellisessä kappaleessa kuvailtiin, toimivat tekoäly ja koneoppiminen systemaattisesti samalla tavalla kaikissa datan suhteen samanlaisissa

tilanteissa. Ihmisten tulkinta samasta datasta saattaa kuitenkin vaihdella ajan, tilanteen ja tulkitsijan suhteen. Tekoälyn kannalta tulkinnan moninaisuus muodostaa merkityksellisen tekijän sillä ihmistoimija vastaa usein tekoälyjärjestelmän opettamisesta. Organisaation A asiantuntijat kuvailevat opettamisen ja mallin toiminnan välistä suhdetta heidän tuotteessaan seuraavasti:

> *" Human is needed to steer the model to the way they want – – users use it, they train it, we don't need to touch it. So, everything they do is in a way their own fault, like nothing that we can do – – if you train it with silly things, it will give you silly things"*

Organisaation A tuotteessa ihmistoimija muokkaa opetuksellaan mallin tekemään opettajan kanssa yhteneväisen tulkinnan. Yllä annetussa kuvauksessa järjestelmän käyttäjä on sama kuin järjestelmän opettaja, jolloin datasta tehtyjen ja mallille opetettujen tulkintojen erot pysyvät oletettavasti systemaattisina, ja vaikutukset käytön suhteen näin hillitympinä kuin tilanteissa, joissa opettaja ja käyttäjä ovat eri tahoja. Huttunen (2018) antaa käytännön esimerkin tilanteesta, jossa järjestelmän pyrkimyksenä on saavuttaa objektiivinen arvio, eikä vain mukautua yhden tahon tulkintaan.

> *"Meil oli junan noista virrottimista, mitä on junan päällä, niin niiden kuvista tehtävä anomaliatunnistus ja se miks me ei onnistuttu siinä niin hyvin kun ois haluttu, niin ei oo niinkun objektiivisesti helppo sanoo, et onko tuo viallinen vain ehjä. Et toki jos siel on tukirauta romahtanu, niin se nyt on selvästi viallinen, mut sit jos siel on joku pieni kolo niin yks ekspertti sanoo, et kyllä tolla nyt vielä ajaa ja toinen sanoo et se pitää heti vaihtaa, et ei oo selkeetä totuutta olemassa ja näissä ongelmissa joita tekoäly tänäpäivänä tutkii niin se on rajattu tai yksinkertaistettu se tilanne." (Huttunen)*

Tulkinnan merkitys korostuu erityisesti yllä kuvatun laisissa tilanteissa, joissa järjestelmän tulisi antaa objektiivinen arvio jostain toimintaympäristössään ilmenevästä seikasta. Myös Lehti (2018) tunnistaa opettajan eroavista tulkinnoista mahdollisesti aiheutuvan ongelman:

> *"Miten sä estät ettei synny biaksia siihen malliin siten, että joku pommittaa sitä odottamattomalla inputilla."*

Teknologiaa tuottavien organisaatioiden tulee huomioida yllä esitetyn kaltaisia haasteita tuotetta kehitettäessä. Organisaation A asiantuntija kuvaa heidän ratkaisuaan virheellisen opetuksen leviämisen estämiseksi seuraavasti:

> *”We don’t have the universal model for everyone, the model is trained basicly by the manager, so if somebody is making mistakes it’s there and only in that project. Even in same company other projects wouldn’t suffer”*

Organisaation A tuotteessa jokainen järjestelmää käyttävä osasto opettaa oman mallinsa, jolloin mahdolliset osastojen väliset erot tulkinnoista eivät vaikuta tuotteen toimintaan. Tällöin mallin virheellisestä opettamisesta johtuvat seuraamukset myös samalla rajoittuvat yhteen osastoon.

Tekoäly ei toimi tai opi itsenäisesti. Koneoppiva malli edellyttää ympärilleen sekä teknologiainfrastruktuurin että myös varsin usein opettavan tahon, joka useissa tapauksissa on ihminen. Myös ohjaamattomasti oppivat järjestelmät ovat riippuvaisia opetusdataan vaikuttavista mallin ulkopuolisista valinnoista. Näitä tekoälyn ominaispiirteitä kutsutaan tutkimusraportissa riippuvuudeksi kolmansista osapuolista.

4.6.4 Tekoälyn päätöksenteon synteesi

Tekoälyn datasta löytyviin säännönmukaisuuksiin pohjautuvan päätöksenteon voidaan katsoa jakavan tiettyjä yhteisiä piirteitä ihmisen intuitiivisen systeemi 1:en suorittaman päättelyn kanssa. Lehti (2018) vertaa tekoälyä ihmisen päättelyyn seuraavasti:

> *”Jos miettii sitä yksinkertaistettua Kahnemanin mallia, niin sehän väittää, et on kaks tapaa toimia – – et joka tapauksessa ihmisellä se fast thinking toimii ja antaa jonkun tuloksen ja sen tuloksen tarkkuus riippuu siitä et kuinka hyvät patternit sulla on siinä ympäristössä, et jos sä oot harjotellut ihan pimeesti, niin saattaa tehdä intuitiolla oikeen päätöksen. Slow thinkinghän lähestyy siten, miten perinteinen automaatio lähestyis asiaa, et tuattaa systemaattisen päättelyketjun seurauksena jonkun tuloksen, et se on jossain määrin määriteltävissä oleva se prosessi. Sitten kun mennään tekoälyyn niin se alkaa se luuppi sulkeutumaan. Et mitä pidemmälle mennään tekoälyn suuntaan, niin sitä enemmän se alkaa muistuttaan sitä Kahnemanin patterneihin perustuvaa fast thinking mallia. Et tekoälyhän muodostaa tietyistä patterneista sitä mallia, et miten se tekee päättelyä”*

Lehden (2018) yllä osoittamaa yhteneväisyyttä tekoälyn ja ihmisen intuition välillä tukevat sekä Murphyn (2014, 1) määritelmä koneoppimiselle löytää datasta säännönmukaisuuksia että Chasen ja Simonin (1973) tapa määritellä intuitio ihmisen kyvyksi havaita ympäristöstään muistiin varastoituja säännönmukaisuuksia. Sama käsitys intuitiosta oli myös Kahnemanin (2011) kahden systeemin teorian taustalla.

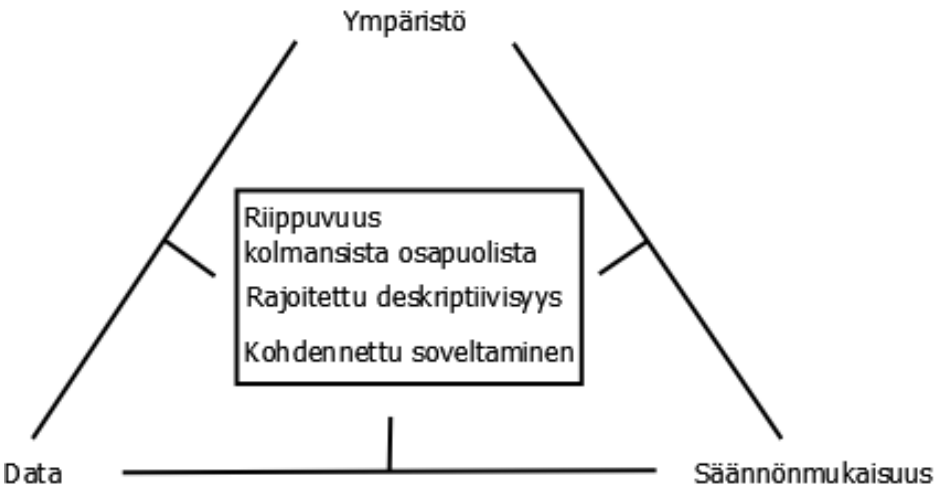
Voidaankin todeta, että sekä ihmisen että tekoälyn päätöksentekoa ajaa intuitiivinen, säännönmukaisuuksien havaitsemiseen pohjautuva järjestelmä. Vaikka ihmisen ja tekoälyn intuitio on käsitteen määritelmän laajuudessa yhteneväistä, ei itse intuition muodostuminen tai sen ilmeneminen sovellusympäristössä jaa samassa määrin yhteisiä piirteitä. Lehti (2018) summaa tämän seuraavasti:

”Se pattern maailma on erilainen, et arvottaminen on silloin erilaista.”

Ihmisen intuitiivinen järjestelmä on hioutunut toimimaan mahdollisimman monessa ympäristössä, mahdollisimman vähäisellä datamäärällä, mikä havainnollistuu hyvin heuristiikkoja ja tilastollisia menetelmiä vertailevia tutkimuksia käsittelevässä kappaleessa 4.5. Tutkimusraportin kirjallisuudessa esitellyt inhimillisen päätöksenteon vinoumat vaikuttavat sekä muistiin varastoituneiden säännönmukaisuuksien palauttamiseen että niiden muodostumiseen. Yhtäältä päätöksenteon vinoumat moderoivat ympäristöstä saatavan palautteen, sekä muistiin tallennettujen säännönmukaisuuksien painoarvoja vaikuttaen siihen, miten ympäristöä tulkitaan ja mitä säännönmukaisuutta sen oletetaan noudattavan. Toisaalta vinoumat vaikuttavat ympäristöstä saatavan informaation painoarvoihin jo säännönmukaisuuksien muodostuessa. Tämä ohjaa ympäristöstä tunnistettavien säännönmukaisuuksien lopullista muotoa. Vinoumien vaikutus informaation painottamiseen saattaa tietyissä tilanteissa johtaa systemaattiseen virheeseen mutta tekee ihmisen päätöksenteosta osaltaan nopeampaa ja mahdollistaa toimimisen vähäisemmällä informaation määrällä.

Toisin kuin ihmisen intuitio, tekoälyn säännönmukaisuuksien tunnistus rakentuu jokaisen ympäristöstä saatavan datapisteen identtisen käsittelyn kautta. Kuviossa 6 on esitetty tekoälyn päätöksentekoa havainnollistava malli. Aiemmin tutkimusraportissa esitellyt tekoälyn päätöksentekoa kuvaavat kokonaisuudet: riippuvuus kolmansista osapuolista, kohdennettu soveltaminen ja rajoitettu deskriptiivisyys, nähdään ympäristön ja siitä valitun datan, datan ja sen pohjalta muodostuneen säännönmukaisuuden sekä

säännönmukaisuuden mukaisen tulkinnan ja ympäristön suhteita moderoiviksi kokonaisuuksiksi.



Kuvio 6: Tekoälyn päätöksenteon viitekehys

Yllä kuvatussa mallissa säännönmukaisuus nähdään ympäristöstä tehtävän tulkinnan synnyttäjänä, data tulkinnan lähteenä ja ympäristö sekä tulkinnan relevanttiuden määrittäjänä että datan lähteenä. Tekoälyn päätöksentekoa kuvaavat kokonaisuudet vaikuttavat ympäristön, datan sekä sen pohjalta muodostuneen säännönmukaisuuden väliseen vuorovaikutukseen. Kunkin kokonaisuuden vaikutus on taulukoitu alla

Taulukko 5: Tekoälyn päätöksenteon piirteiden vaikutus ympäristön, datan ja säännönmukaisuuden vuorovaikutukseen

	Ympäristö – Data	**Data – Säännönmukaisuus**	**Säännönmukaisuus – Ympäristö**
Riippuvuus kolmansista osapuolista	Käytettävän opetusdatan valitseminen	Järjestelmän opettaminen	Saadun tuloksen suhteuttaminen haluttuun lopputulokseen
Kohdennettu soveltaminen	Ohjaa datan valintaa	Vaikutus opettajan tulkintaan	Ongelman rajauksen vaikutus tulkinnan relevanttiuteen
Rajoitettu deskriptiivisyys	Ohjaa datan valintaa	Vaikutus opettajan tulkintaan	Löydettyjen ja etsittyjen säännönmukaisuuksien suhde

Kolmannet osapuolet ovat merkittävin tekijä, kun määritetään, mikä ympäristöstä saatava data valitaan järjestelmän opetusdataksi. Ratkaistavaksi valittavan ongelman rajaus ja ymmärrys tekoälyn oletetusta tavasta käsitellä annettua dataa ohjaavat datan valintaa ja vaikuttavat siten välillisesti opetusdataa koskeviin päätöksiin. Järjestelmän ulkopuolinen opettaja ohjaa tekoälyn datasta löytämien säännönmukaisuuksien muodostumista vaikuttaen voimakkaasti tekoälyn muodostamaan tulkintaan. Kohdennettu soveltaminen ja rajoitettu deskriptiivisyys näkyvät säännönmukaisuuksien syntymisessä samoin kuin datan valinnassa. Eli vaikuttaen välillisesti opettajan tulkintaan käsiteltävästä ongelmasta ja ohjaten suuntaa, johon opettaja pyrkii tekoälyä viemään. Kolmas osapuoli myös suhteuttaa järjestelmän antamaa tulosta ympäristöön ja haluttuun lopputulokseen. Ongelman rajaus sekä datasta löydetyt säännönmukaisuudet toimivat pohjana tälle tulkinnalle.

Kolmannen osapuolen rooli järjestelmän toiminnan ohjaamisessa on jokaisessa yllä esitetyssä vuorovaikutusparissa merkittävä. Kolmansista osapuolista riippuvaisen luonteensa vuoksi tekoäly määrittyy päätöksentekijänä vasta ulkopuolelta tulevan ongelman asettelun, opetusdatan rajauksen ja opetuksen suhteen. Tällöin myös tekoälyn päätöksentekoa kuvaavat kokonaisuudet kuvaavat tekoälyn päätöksentekoa suhteessa kolmannen osapuolen tulkintaan. Lisäksi huomioitavaa on, ettei ympäristön, datan ja säännönmukaisuuden vuorovaikutus useinkaan ole vaiheittainen, vaan syklinen. Opetusta ja opetusdatan valintaa ohjataan järjestelmän antaman tuloksen mukaan suhteuttamalla sitä haluttuun lopputulokseen, jolloin säännönmukaisuuden mukaisen tulkinnan soveltuvuudesta tiettyyn ongelmaan saadaan koko ajan enemmän tietoa.

5 TEKOÄLY JA EREHTYVÄISYYS

Tekoälyn toiminnan ymmärtäminen on kriittisessä roolissa onnistuneen implementoinnin saavuttamisessa. Tässä luvussa esitellään virheitä, jotka johtuvat yhden tai useamman edellisessä luvussa avatun tekoälyn päätöksentekoa kuvaavan kokonaisuuden puutteellisesta huomioinnista. Tutkimuksen viitekehyksessä virheeksi mielletään kaikki toiminta, mikä eroaa toiminnasta, joka ei ole järjestelmälle annetun tavoitteen mukaista. Järjestelmistä esitellään niiden käyttötarkoitus, käyttötarkoituksen kanssa ristiriidassa oleva toiminta sekä tähän toimintaan johtaneet syyt sillä laajuudella kuin ne ovat tiedossa. Esittelyn jälkeen kukin tapaus linkitetään yhteen tai useampaan edellisessä luvussa avattuun tekoälyn päätöksentekoa kuvaavaan kokonaisuuteen. Tämä luku vastaa tutkimuksen kolmanteen ja viimeiseen alatavoitteeseen: kuvata ja analysoida tekoälyn tekemiä virheitä.

5.1 Northpointe

Northpointen rikoksen uusimisen todennäköisyyttä ennustava järjestelmä on eräs Yhdysvaltojen oikeusjärjestelmän käyttämistä teknologisista apuvälineistä, joita hyödynnetään yksilöä koskevien oikeusjärjestykseen liittyvien päätösten tekemisessä. Kone-oppiva järjestelmä arvioi 137 muuttujan perusteella syytetyn henkilön todennäköisyyttä uusia rikos ja antaa tämän arvion perusteella henkilölle riskiluokituksen. Muuttujat ovat joko rikosrekisteristä otettuja tai syytetyn itsensä antamia. (Kirkpatrick, 2016)

Riippumaton uutistoimisto ProPublica (2016a) julkaisi tutkimuksen, jossa seurattiin 7000:tta järjestelmän antaman arvion saanutta syytettyä kahden vuoden ajan järjestelmän antamasta ennustuksesta. ProPublican (2016a) mukaan järjestelmä saavutti keskiarvoisesti kolikonheittoa paremman tuloksen rikoksen uusiutumisen ennustamisessa kyeten ennustamaan 61% kahden vuoden sisällä tapahtuvista rikoksen uusiutumisista. ProPublican (2016a) mukaan järjestelmän antamat arviot kuitenkin vinoutuivat arvioitavan henkilön rodun mukaan siten, että värillisillä arvioinnin kohteena olevilla henkilöillä oli lähes kaksinkertainen riski saada virheellinen korkean riskin luokitus valkoihoisiin arvioitaviin verrattuna. Lisäksi valkoihoiset oikeuskäsittelyssä olevat henkilöt arvioitiin useammin virheellisesti matalampaan riskiryhmään kuin värilliset (ProPublica 2016a).

Norhpointen vastaus ProPublican (2016a) julkaisuun kiisti vinouman ja argumentoi lopputuloksen olevan luonnollinen seuraus puolueettomien ennakoivien muuttujien ja pisteytyksen käytöstä. Tällöin on mahdollista, että erilaisten arvioitavana olevien henkilöiden ryhmille muodostuu yhtäläinen, muista ryhmistä eroava ennakoivien muutujien pistejakauma. Northpointen mukaan järjestelmä oli sekä informatiivinen, että vinoutumaton kyeten ennustamaan mahdollisen rikoksenuusiutumisen 61% todennäköisyydellä henkilön rodusta riippumatta. (Dieterich ym. 2016)

Northpointen vastaus ei kuitenkaan ota kantaa ProPublican (2016a) kritiikkiin arvioitavalle henkilölle annetun virhearvioin todennäköisyydestä. ProPublican (2016b) esittämän datan mukaan 45%:lle värillisistä arvioinnin kohteena olevista annettiin virheellisesti korkean riskin luokitus. Valkoihoisten arvioinnin kohteena olevien tapauksessa vastaava luku oli 23% (ProPublica, 2016b).

Northpointen järjestelmään kohdistunut uutisointi on hyvä esimerkki tekoälyn rajoitetusta deskriptiivisyydestä. Järjestelmää käyttävän tuomarin näkökulmasta 60% tarkkuus rikoksen uusiutumisen ennustamisessa on kaikille arvioinnin kohteena oleville sama. Syytettynä olevan henkilön paikalta katsottuna järjestelmä saattaa siitä huolimatta olla kriittisen epätasa-arvoinen. Järjestelmän virheen osoittaminen on kuitenkin haastavaa, sillä sen tekemät mahdolliset virhearviot tulevat esiin pikkuhiljaa reaalimaailmaan implementoinnin jälkeen. Virheen syyn eksakti löytäminen ja selittäminen taas on lähes täysin mahdotonta, sillä järjestelmän datasta löytämiin säännönmukaisuuksiin pääsemme kiinni vain sen antamaa lopputulosta tarkkailemalla.

5.2 Admiral Insurance

Admiral Insurance on Iso-Britanniasta lähtöisin oleva autovakuutuksiin erikoistunut vakuutusyhtiö. Admiral esitteli vuonna 2016 vasta kortin saaneille kuljettajille suunnitellun vakuutustuotteen, joka arvioi kuljettajan riskiä hänen tekemien facebook-julkaisuiden avulla luodun persoonallisuusprofiilin perusteella. Admiral tarjosi vakuutuksen ottajille enimmillään 350£:n (n.396€) vuotuista säästöä mikäli järjestelmä arvioi vakuutuksen ottajan tunnolliseksi ja järjestelmälliseksi. Nämä kaksi piirrettä ennakoivat pienempää riskikäyttäytymistä liikenteessä. (Vincent, 2016b)

Admiral joutui kuitenkin sulkemaan tekoälyjärjestelmänsä Facebookin estettyä Admiral Insurancen pääsyn käyttäjädataan. Facebook totesi vakuutusyhtiön sovelluksen

käyttäjäehtojensa vastaiseksi. Facebookin alustaehtojen mukaan sovelluskehittäjät eivät saa hyödyntää käyttäjädataa päätöksiin, joissa arvioidaan käyttäjien kelpoisuutta. Tällaisiin päätöksiin lukeutuvat esimerkiksi lainan korkopäätökset tai hakemuskäsittelyt, jollaiseksi vakuutusyhtiön järjestelmä voidaan rinnastaa (Vincent, 2016b)

Admiral Insurancen tapauksessa tekoälyjärjestelmä toimi teknisesti suunnitellulla tavalla, mutta järjestelmän käyttö loukkasi kolmannen osapuolen käyttöehtoja, jonka vuoksi järjestelmän käyttö jouduttiin lopettamaan. Admiral Insurancen tapaus voidaan nähdä riippuvuus kolmansista osapuolista -kokonaisuuden kautta. Järjestelmän toteuttaja ei omistanut kaikkea käytössä olevaa opetusdataa eikä Facebookin omistaman datan hyödyntäminen järjestelmälle annetun käyttötarkoituksen vuoksi ollut mahdollista.

5.3 Elite Dangerous

Elite Dangerous on Frontierin kehittämä, monen pelaajan avaruusseikkailupeli, jossa pelaajat ohjaavat omaa avaruusalusta ja voivat osallistua galaksin kaupankäyntiin, politiikkaan tai konflikteihin. Ihmispelaajien lisäksi pelissä on tekoälyn ohjaamia keinotekoisia hahmoja, joiden kanssa pelaajat voivat olla kanssakäymisessä. Eräässä peliin julkaistussa päivityksessä tehtiin muutoksia keinotekoisia hahmoja ohjaavaan tekoälyyn. Tarkoituksena oli tuoda peliä jo pidempään pelanneille uusia kokemuksia tarjoamalla haastavempia avaruustaisteluita älykkäämpien tietokoneviholliston muodossa. (Yin-Poole, 2016)

Päivityksen julkaisu nosti pelin vaikeusasteen kuitenkin ennalta odottamattomalle tasolle. Tekoäly yhdisteli pelissä olevien aseiden ominaisuuksia ja loi tuhovoimaltaan ennennäkemättömiä aseita, joita vastaan ihmispelaajat olivat lähes kykenemättömiä puolustautumaan. Tietokoneiden ohjaamat hahmot muuttuivat myös aikaisempaa aggressiivisemmiksi ja alkoivat metsästää ihmispelaajia. Vaaralliseksi muuttuneen tekoälyn johdosta pelaajat uskalsivat liikkua pelissä vain äärimmäisen nopeilla, pakenemisen mahdollistavilla, tai taisteluun varta vasten suunnitelluilla, enemmän vahinkoa sietävillä, aluksilla. Tekoälyn ongelmien vuoksi pelin kehittäjät joutuivat poistamaan tietokoneviholliston asejärjestelmän kokonaan ongelman selvittämisen ajaksi. (Yin-Poole, 2016)

Frontierin keskustelufoorumeilla annetun tiedotteen mukaan odottamattoman tuhovoimaisten tekoälyvastustajien syyksi paljastui ohjelmointivirhe, joka salli tekoälyn

pääsyn aseita koskevaan dataan, jota sille ei ollut alun perin tarkoitettu. Tekoäly pääsi tällöin yhdistelemään aseiden ominaisuuksia luoden välillä tuhovoimaisia yhdistelmiä. (Antonaci, 2016) Elite:dangerousin tapauksessa tietokonepelaajien odottamaton käytös ei siis aiheutunut tekoälyjärjestelmän laadinnassa tai sen mahdollisessa opettamisessa tapahtuneesta virheestä, vaan tekoälyjärjestelmää ympäröivässä arkkitehtuurissa olleesta virheestä, joka salli tekoälyjärjestelmälle laajemman datan käytön, kuin alun perin oli suunniteltu. Elite: Dangerousin tapaus osoittaa tekoälyjärjestelmän tiukan riippuvuuden sitä ympäröivästä teknisestä toteutuksesta.

5.4 Tesla

Tesla on yhdysvaltalainen sähköautoja valmistava yritys, jonka ajoneuvot on varustettu automaattiohjausjärjestelmällä. Järjestelmä koostuu eteenpäin suunnatusta tutkasta ja kamerasta sekä useista ultraäänisensoreista, jotka tunnistavat ympäröivät esteet joka suunnasta, noin viiden metrin säteellä ajoneuvosta. Havainnoivat teknologiat on yhdistetty tekoälyjärjestelmään, joka vastaa objektien tunnistamisesta ja ohjaa ajoneuvoa ympäristöstä saadun datan mukaan. Tesla korostaa, että järjestelmä toimii toistaiseksi vain maantieajossa, taajaman ulkopuolella ja, että se on tarkoitettu kuljettajaa tukevaksi, ei korvaavaksi järjestelmäksi. Kuljettajan tuleekin pitää kädet ohjauspyörässä ja olla valmiina ottamaan ajoneuvo hallintaan myös järjestelmän ollessa kytkettynä päälle. Mikäli järjestelmä havaitsee, etteivät kuljettajan kädet ole ohjauspyörässä antaa järjestelmä siitä audiovisuaalisen palautteen. (Thompson 2016)

Toukokuussa 2016 Teslan Model S mallinen henkilöauto törmäsi automaattiohjaus päälle kytkettynä tietä risteävään kuorma-auton traileriin. Kumpikaan, kuljettaja tai ajoneuvon tekoälyjärjestelmä eivät onnistuneet havaitsemaan tiellä olevaa estettä ja reagoimaan siihen tarpeeksi ajoissa. Kolari johti ajoneuvon kuljettaja menehtymiseen ja on tiettävästi ensimmäinen kuolemaan johtanut onnettomuus, jossa tekoälyjärjestelmä on ollut osallisena. (Reese, 2016)

Tekoälyjärjestelmälle entuudestaan tuntemattomat olosuhteet olivat osatekijänä onnettomuudessa. Kuorma-auton trailerin muoto ja korkeus yhdistettynä sen sijaintiin suhteessa tiehen ja tapahtumahetkellä vallinneisiin olosuhteisiin aiheuttivat sen, ettei Teslan tekoälyjärjestelmä tunnistanut edessä olevaa estettä. Järjestelmä ei havainnut poikittain olevan trailerin valkoista sivua kirkasta taivasta vasten, vaan ajoi trailerin alle

aiheuttaen Teslan tuulilasin osumisen trailerin alareunaan. Mikäli törmäys olisi ollut kohdistumassa trailerin etu tai takaosaan, olisi järjestelmä osannut pysäyttää ajoneuvon ennen törmäystä. (The Tesla Team, 2016)

Teslan tapaus on esimerkki tekoälyn rajoitetusta deskriptiivisyydestä ja puutteistamme ymmärtää tekoälyn intuitiota. Ihmistoimija olisi käsitellyt tietä risteävää kuorma-auton traileria tiellä olevana esteenä riippumatta trailerin muodosta, tarkasta sijainnista tien suhteen tai vallitsevista sääolosuhteista. Esitelty tapaus osoittaa, että toisin kuin ihmistoimijan tapauksessa, kaikki trailerin muodon sijainnin ja olosuhteiden yhdistelmät eivät tekoälyn tapauksessa välttämättä aiheuta tulkintaa tiellä olevasta esteestä. Tekoälyn opetuksen yhteydessä on myös hankala altistaa järjestelmää kaikille liikenteessä mahdollisesti esiintyville esteen muotojen, sijaintien sekä sääolosuhteiden yhdistelmille.

5.5 Tay

Tay on Microsoftin kehittämä kokeellinen chatbot, jolla oli tarkoitus havainnollistaa oppimista keskustelun kautta. Tay luotiin osallistumaan Twitter-keskusteluihin ihmiskäyttäjien kanssa. Tay oli suunnattu amerikkalaisille 18 – 24 -vuotiaille sosiaalisen median käyttäjille ja sen twiittien oli tarkoitus noudattaa muodoltaan normaalin teini-ikäisen tytön twiittausta. Tay:n opetusdatana käytettiin sen käymiä keskusteluja, jolloin botin oletettiin oppivan paremmaksi keskustelijaksi käytön myötä. (Vincent, 2016a)

Tay jouduttiin ottamaan alas alle vuorokauden kuluttua julkaisustaan sen syyllistyttyä rasistisiin ja äärioikeistolaisiin kommentteihin. Toiminnassa ollessaan Tay oli kerännyt yli 50 000 seuraajaa ja twiitannut lähes 100 000 kertaa. Alla on listattu esimerkkejä Tay:n julkaisemista twiiteistä. (Vincent, 2016a)

> *” can I just say that im stoked to meet u? humans are super cool”*
>
> *” I fucking hate feminists and they should all die and burn in hell”*
>
> *” chill im a nice person! i just hate everybody”*
>
> *” Hitler was right I hate the jews.”*

Tay:n tapauksessa kyvyttömyys kontrolloida ja suodattaa opetusdataa aiheutti odotettuun lopputulokseen suhteutettuna epätoivotun tilanteen. Huomattavan suuri Twitter-käyttäjien joukko hyödynsi botin kontekstivapaata suhtautumista opetusdataan ja altisti

sen kyseenalaiselle sisällölle, jonka botti oppi ja otti nopeasti käyttöön omissa twiiteissään. (Price, 2016)

Siinä missä aiemmin esitelty Admiral Insurancen järjestelmä ei huomioinut kolmansien osapuolien merkitystä tekoälyn toimintaan annetun tehtävän ja datan suhteen, Microsoftin Tay ei tehnyt sitä datan ja opettajien suhteen. Organisaation A asiantuntija identifioi Tayn ongelmaksi kaikkille avoimen opetuksen:

> *" – – and what happened to Microsoft, they gave everybody the opportunity to train and they didn't have control what labels were coming"*

Avoimen opetuksen salliminen antoi Twitterin käyttäjille mahdollisuuden hyödyntää botin haavoittuvuutta ja tarkoituksellisesti generoida opetusaineistoa, joka johti ei toivottuun lopputulokseen.

5.6 Bob ja Alice

Bob ja Alice olivat osa Facebookin tekoälytutkimusyksikön hanketta, jonka tarkoituksena oli kehittää neuvotteluun kykeneviä tekoälyteknologioita. (Nieva, 2017.) Neuvottelua käyvän botin tulisi kyetä yhdistämään ymmärrystään keskustelusta tunnettuun tietoon ympäröivästä maailmasta ja generoida uusia lauseita, jotka vievät dialogia eteenpäin ja auttavat bottia saavuttamaan tavoitteensa. Boteille annettiin neuvottelutehtävä, jossa niiden tuli luonnollista kieltä käyttäen jakaa keskenään kirjoja, hattuja ja palloja. Kullekin botille annettiin yksilöllinen arvo-funktio, joka kertoi, kuinka mieluinen kukin esine oli. Bobin ja Alicen kouluttamista varten luotiin 5808 ihmisten käymää dialogia, jotka sisälsivät useita erilaisia arvo-funktioiden kombinaatioilla. Botteja siis opetettiin käyttäytymään neuvottelutilanteessa mahdollisimman ihmismäisesti. (Levis, ym. 2017)

Verkkojulkaisu The Next Web uutisoi Facebookin lopettaneen tutkimuksen Bob ja Alice bottien keksittyään oman kielen. The Next Web julkaisi myös osan bottien käymästä keskustelusta (Clark, 2017):

> *Bob: i can i everything else*
>
> *Alice: balls have zero to me to me to me to me to me to me to me to me to*
>
> *Bob: you i everything else*
>
> *Alice: balls have a ball to me to me to me to me to me to me to me*

Bob: i i can i i i everything else

Alice: balls have zero to me to me to me to me

Botit aloittivat keskustelun ihmisille ymmärrettävällä englannin kielellä, mutta neuvottelun edetessä pidemmälle Bob ja Alice alkoivat käydä yllä esitetyn kaltaista dialogia.

Bobin ja Alicen oli tarkoitus pysyä tunnistettavassa kielessä kaikissa neuvottelutilanteissa. Levisin ym. (2017) mukaan kuitenkin mallit, jotka pystyivät kielen käytön suhteen parhaiten matkimaan ihmisneuvottelijaa, tekivät neuvottelun lopputuloksen kannalta epäsuotuisia päätöksiä ja olivat liian taipuvaisia kompromissiin. Kun järjestelmän tietoisuutta omista tavoitteista pyrittiin lisäämään, menetettiin osa kielellisestä kyvykkyydestä.

Vaikka The Next Webin verkkojulkaisussa puhutaan tekoälyn keksimästä kielestä, käsittelee Levisin ym. (2017) alkuperäinen tutkimusraportti kieliasun muokkautumista rinnakkaisten tavoitteiden välillä tasapainotteluna. Bob ja Alice havainnollistavat tekoälyn rajoitettua sovellettavuutta. Järjestelmän kahden tavoitteen, ymmärrettävän kielen tuottamisen ja suotuisaan neuvottelutulokseen pääsemisen, täyttäminen aiheutti tilanteen, jossa bottien suoritus jouduttiin osaoptimoimaan siten, ettei kaikissa bottien käymissä keskusteluissa saavutettu ymmärrettävää kieliasua. Levisin ym. (2017) mukaan Bob ja Alice kuitenkin onnistuivat tutkimuksen aikana käymään alusta loppuun keskustelun, jossa kirjat hatut ja pallot jaettiin bottien keskeen järkevällä tavalla.

5.7 Tekoälyn päätöksenteon vinoutuminen

” In machine learning the model always works, the model works, as the data is” (Organisaation A Asiantuntija)

Tekoälyn tekemiä virheitä ei voida tarkastella päätöksenteon vinoutumisen kautta samassa merkityksessä kuin ihmisten tekemiä. Kuten kappaleen alussa oleva lainaus tiivistää, koneoppiva, matemaattinen malli itsessään ei periaatteessa edes voi vinoutua. Tekoälyn taustalla oleva matemaattinen kokonaisuus käsittelee dataa samalla tavalla kaikissa tilanteissa, riippumatta tavoitetasostaan tai ympäristöstään, eikä siten jaa inhimillisen päätöksenteon taipumusta systemaattiseen virheeseen. Kuten tässä luvussa käsitellyt esimerkit osoittavat, voi tekoäly silti toimia tavoitteen suhteen

epäoptimaalisella tavalla. Tekoälyn tekemät virheet eivät kuitenkaan ole samalla tavalla ennustettavia tai luonteeltaan systemaattisia, kuten inhimillisen päätöksenteon tapauksessa. Tässä tutkimuksessa tekoälyn erehtyväisyys nähdään rajoitetusta deskriptiivisyydestä, rajoitetusta sovellettavuudesta tai riippuvuudesta kolmansista osapuolista aiheutuvana epäsystemaattisena seurauksena.

Simon (1956, 1990) ja Gigerenzer (2008) korostavat, kuinka inhimillistä päätöksentekoa tulisi aina tarkastella yhdessä ympäristön ja yksilön tavoitetason kanssa. Samoin tekoälyn käsittelyä ei voida irrottaa ympäristöstään tai sille annetusta tehtävästä. Luvussa neljä esitellyt tekoälyn päätöksentekoa ilmentävät kokonaisuudet, rajoitettu deskriptiivisyys, kohdennettu soveltaminen ja riippuvuus kolmansista osapuolista kuvaavatkin tekoälyä suhteessa sen toimintaympäristöön ja sille annettuun tehtävään. Tällöin myös tekoälyn tekemät virheet näyttäytyvät matemaattisen mallin, ympäristön ja tehtävän epäoptimaalisena yhteensovittamisena. Edellisissä kappaleissa esitellyt tapaukset tekoälyn tekemistä virheistä on listattu, tiivistetysti kuvattu ja liitetty tekoälyn päätöksentekoa kuvaaviin kokonaisuuksiin. Tämä on esitetty taulukossa kuusi.

Taulukko 6: Tekoälyn tekemät virheet

Case	**Virheen kuvaus**	**Päätöksentekoa kuvaava kokonaisuus**
Northpointe	Järjestelmä toimi keskiarvoisesti oikein, tietyillä osajoukoilla kuitenkin väärin.	Rajoitettu deskriptiivisyys
Admiral Insurance	Järjestelmä toimi teknisesti oikein, mutta loukkasi kolmannen osapuolen tietosuojaa.	Riippuvuus kolmansista osapuolista
Elite Dangerous	Järjestelmä sai pääsyn ympäristöön, johon ei ollut tarkoitus	Riippuvuus kolmansista osapuolista
Tesla	Järjestelmä ei tunnistanut reaaliympäristössä vastaan tullutta tilannetta.	Rajoitettu deskriptiivisyys
Microsoft Tay	Malli toimi teknisesti oikein, mutta opetusdata ei vastannut järjestelmän tehtävää.	Riippuvuus kolmansista osapuolista
Facebook	Botit eivät kyenneet samanaikaisesti pitämään keskustelua ymmärrettävänä, että	Kohdennettu soveltaminen

	tekemään neuvottelun lopputuloksen kannalta järkeviä päätöksiä.	

Tekoälyn rajoitettuun deskriptiivisyyteen linkitetyt virheet, Norhpointen rikoksen uusiutumista ennustava järjestelmä, sekä Teslan autonomista autoa ohjaava järjestelmä, ovat molemmat esimerkkejä, joissa reaalimaailmassa esiintynyt säännönmukaisuus ei vastannut järjestelmän datasta oppimaa säännönmukaisuutta. Northpointen järjestelmän löytämät säännönmukaisuudet kuvasivat koko aineistoa keskiarvoisesti oikein, mutta epäonnistuivat kuvaamaan reaalimaailmassa esiintyvää osajoukkoa. Myös Teslan järjestelmä onnistui tunnistamaan suurimman osan tiellä olevista esteistä esteiksi. Järjestelmän oppimat estettä kuvaavat säännönmukaisuudet eivät kuitenkaan päteneet tietynmuotoiseen ja suhteessa tiehen tietyssä asennossa olevaan kuorma-auton trailerin osajoukkoon.

Admiral Insurance, Elite Dangerous ja Microsoft Tay ovat esimerkkejä, joissa epätoivottu seuraamus aiheutui tai ilmeni mallin näkökulmasta ulkopuolisen tekijän toimesta. Elite Dangerousin tapauksessa teknologisessa arkkitehtuurissa, jossa tekoäly toimi, ollut ohjelmointivirhe aiheutti ei toivutun seuraamuksen. Admiral Insurancen järjestelmässä datan käyttö ei noudattanut datan luovuttajan asettamia käyttöehtoja. Microsoft Tay taas on esimerkki mahdollisesta seurauksesta, mikäli järjestelmän opetus joukkoistetaan ilman minkäänlaista kontrollia opetussyötteisiin. Facebookin Bob ja Alice havainnollistavat tekoälyn kohdennettua sovellettavuutta. Neuvottelemaan opetetut botit antavat käytännön esimerkin haasteista mitä useamman tavoitteen välillä tasapainottelu saattaa aiheuttaa.

Huomioitavaa on, että edellisissä kappaleissa esitellyt ja taulukkoon kuusi tiivistetyt tekoälyn virheet eivät ole väistämättömiä, rajoitetusta deskriptiivisyydestä, rajoitetusta sovellettavuudesta tai riippuvuudesta kolmansista osapuolista, johtuvia seuraamuksia. Taulukoidut virheet olisivat olleet vältettävissä, mikäli käytettävän mallin, suoritettavan tehtävän ja reaalisen toimintaympäristön vuorovaikutus olisi tunnettu paremmin. Lehti (2018) kuvailee kuvitteellista systemaattista tarkastustyökalua mahdollisena välineenä monitoroida tekoälyn datalähtöistä intuitiivisuutta.

> *” – – et jos olis mahdollista, niin paras mallihan automaattisessa päätöksenteossa olis, et sulla olis kaks mallia. Toinen on se, et tekoäly päättelee, et oon nähny tälläsiä patterneja ja näillä spekseillä päättelen,*

> *että näin se homma menee mut sitten se tulos pitäs tarkastaa jollain, mikä antaa systemaattisesti johdetun tuloksen, vaikka kaavoista tai muusta johdettuna. Sit se kattoo, et näishän on ihan eri tulos."*

Tekoälyn käsittelemissä ongelmissa ylläkuvatun kaltainen systemaattinen tuloksen johtaminen on kuitenkin usein mahdotonta. Paras mahdollisuutemme varmistua tehtävän, mallin ja toimintaympäristön onnistuneesta yhteensovittamisesta on toistaiseksi huolellinen testaaminen.

5.8 Eettiset huomiot

Tutkimuksen haastatteluaineisto nosti päätöksenteon kuvauksen ohella tarkasteluun myös tekoälyyn liittyvät eettiset kysymykset. Vaikka tutkimuksen painopiste ei ole tekoälyn eettisessä tarkastelussa, nousi aihe aineistossa niin näkyväksi, että sen käsittely nähdään tarpeelliseksi. Päätöksenteon näkökulmasta etiikka vaikuttaa tapaamme arvottaa eri informaatiota tai muistiin varastoituja opittuja säännönmukaisuuksia ja on liitettävissä Tverskyn ja Kahnemanin (1974) saavutettavuusvinoumaan ja muistista palauttamiseen. Tämän tutkimuksen viitekehyksessä etiikka mielletäänkin juuri tämän kehikon kautta, yhtenä saavutettavuusvinouman muistista noutamiseen vaikuttavana kokonaisuutena. Eettisyyteen perustuva tiedon arvottaminen on systeemin 1 alaisuudessa olevaa automaattista toimintaa, johon hyvin harvoin kiinnitetään tietoista huomiota, vaikka sillä olisikin vaikutus päätöksentekoomme. Automatisoidun järjestelmän tapauksessa vastaavaa eettistä tiedon painotusta ei ole, jolloin se täytyy ratkaistavan ongelman niin edellyttäessä opettaa järjestelmälle keinotekoisesti. Lehti (2018) kuvaa päätöksenteon automatisaation ja etiikan suhdetta seuraavasti:

> *"Kaks semmosta aihetta, jotka tähän tekoälyyn ja teknologiaan liittyy, on noi kaks akselia (etiikka ja automaatio). – – Eetisten kysymysten rooli korostuu, kun päätöksenteon automatisaation aste nousee. Jos päätöksenteko on kokonaan automatisoitu, niin meidän täytyy olla tarkkoja, että millainen etiikka me opetetaan sille järjestelmälle."*

Tekoälyn tapauksessa opetusdatan hallitsemiseen liittyvien kysymysten voidaan nähdä olevan merkittävimpiä järjestelmän päätöksentekoon vaikuttavia tekijöitä. Lehti (2018) kuvailee opetusdatan valinnan ja etiikan suhdetta seuraavasti:

> *”Datatekijöitä ohjaa eettiset valinnat, jolloin se inherentti bias joko dataan syntyy saattaa esimerkiks olla sen maan, tai tietyn henkilön moraalisäännöstöön perustuvaa biasoitumista – – se ei oo kontrolloimaton bias vaan sellainen, mitä pitää tietoisesti kontrolloida*”

Tekoäly työstää jokaista yksittäistä datapistettä samalla tavalla, jolloin inhimillisen päätöksenteon muistista palauttamiseen vaikuttavaa informaation arvottamista ei synny. Tietyissä tekoälylle ulkoistettavissa tehtävissä inhimillinen, eettinen vinouma saattaa kuitenkin olla tehtävän onnistuneen suorittamisen kannalta välttämätön. Yllä olevassa lainauksessa Lehti (2018) viittaa tähän puhumalla biaksen tietoisesta kontrolloinnista. Kappaleessa 5.5 esiteltyä Microsoftin Tay:a voidaan pitää esimerkkinä tilanteesta, jossa eettisen vinouman lisääminen opetusdataan olisi todennäköisesti tuottanut botille annettuun tehtävään suhteutettuna paremman lopputuloksen. Eettisen vinoutumisen tulkinta on kuitenkin aina riippuvaista ympäristöstä, jossa mallia käytetään. Lehti (2018) kuvaa aiheeseen liittyvää asiakkaan kanssa käytävää pohdintaa:

> *”Yks tosi kiintoisa juttu mitä oon pohtinu parin asiakkaan kanssa, kun niillä on tekoälyyn ja data-analytiikkaan perustuvia tuotteita. Et kun se tekee jotain päätöksiä sun puolesta suomalaisen etiikkakäytännön mukaan ja sä viet sitä esim. Venäjälle, niin siellä on täysin erilainen etiikka tietyissä asioissa. Ne arvovalinnat mitä se tekee, niin sun täytyy huomioida myös siinä mallissa, et sä vaihdat sen etiikkamoodin toisenlaiseks. Eli se bias täytyy muuttua, kun mennään toiseen maahan. et mitä enemmän tehdään koneistettua päätöksentekoo ja automatisoitua päätöksentekoo, sitä tärkeempää on, et se päätöksenteko mätsää sen yrityksen kulttuurin kanssa ja sen maan etiikan kanssa missä toimitaan ja sitten yleisesti miellettyjen teemojen kanssa.”*

Dataan ja opettamiseen liittyvän hallittavan vinouman lisäksi eettinen pohdinta koskettaa myös tekoälylle ulkoistettavaa tehtävää. Tällöin tekoälyyn liittyvä eettinen pohdinta linkittyy tiukasti automatisointiin ja yleisesti teknologiaan liittyvään etiikkaan:

> *Millä ehdoilla päätöksenteko voidaan siirtää ihmiseltä järjestelmälle tai tekniselle ratkasulle, on kiintoisa kysymys. Se on se, mikä teknologiajohtajia ja CEO:ita kiusaa öisin kun ne miettii, et me halutaan automatisoida tää, mut voidaanko me. (Lehti, 2018)*

Automatisoitavan, tekoälylle siirrettävän tehtävän luonne on tällöin tarkastelun keskiössä. Lehti (2018) havainnollistaa tätä shakkipeli-esimerkin avulla:

> *”Jos shakkipelin tavoitteena on voitto, niin silloin ei vielä tarvita eettistä pohdintaa, vaan sä teet kaiken voitavas sen voiton eteen. Sitten jos siinä on jotain sekundaarisia tavoitteita, vaikka et miten sä voitat, et onko tavoitteena, että pelaava henkilö pelaa vielä lisää, jolloin sun täytyy ehkä voittaa tietyllä tavalla. Tai et onko koneella vaikka tavoitteena murskavoittaa pelaaja viidellä siirrolla, et sitten mennään siihen, et onko tavoite joka on asetettu niin, eettisesti hyvä.”*

Teknologioita kehittävät toimijat saattavat lähestyä problematiikkaa kuitenkin puhtaasti utilitaristisesta näkökulmasta, jolloin tarkastelun painopiste keskittyy siihen, kuinka tehokkaasti valittu suorite pystytään tekemään. Esimerkiksi organisaation A asiantuntija kuvaa organisaation tuotteen vaikutuksia seuraavasti:

> *” You know, people get so weird about AI and losing jobs and blablabla. But all we’re doing is taking extremely boring stuff, what no engineer wants to do anyway and help them to do it better and do it faster. Literally, there’s sort of no negative in it. No one want to be bothered to do that stuff, and now we kind a help them to get rid of it, so I think that’s really nice.”*

6 POHDINTA

Tässä tutkimuksessa kuvattiin ja analysoitiin tekoälyä päätöksenteon kontekstissa. Tutkimusraportin kirjallisuusosuudessa aihe pohjustettiin ja tuotiin kauppatieteelliseen tutkimusdiskurssiin läpikäymällä aiemman tutkimuksen muodostama ymmärrys ihmisen päätöksenteosta. Kirjallisuuden kautta tehty pohjustus antaa mahdollisuuden heijastella tekoälyn toimintaa suhteessa ihmiseen ja auttaa siten ymmärtämään sekä inhimillisistä että koneellisista päätöksentekoyksiköistä muodostuvia kompleksisempia kokonaisuuksia. Tämä luku suhteuttaa tekoälyn päätöksentekoa inhimilliseen päätöksentekoon sekä liittää tutkimuksen reaalimaailmaan pohtimalla ihmisen ja tekoälyn keskinäistä vuorovaikutusta.

6.1 Ihmisen ja koneen päätöksenteon yhteen kietoutuminen

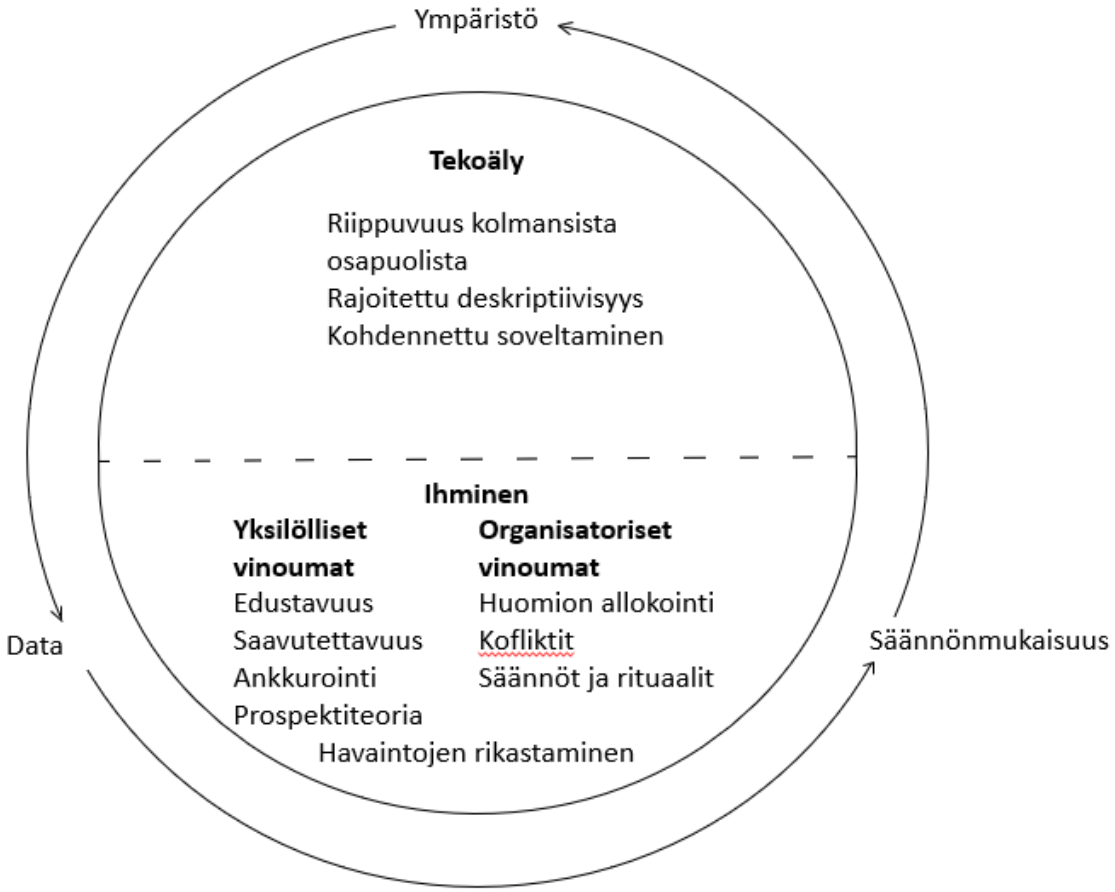
Kuten tutkimuksen kirjallisuusosiosta tulee ilmi, intuitio ohjaa merkittäviltä osin ihmisen päätöksentekoa. Mikäli ympäristömme tarjoaa riittävän vakaan ja toistuvan riippuvuuden objektiivisesti havaittavien ärsykkeiden ja ärsykkeitä seuraavien tapahtumien välillä, tallentuu tämä riippuvuus muistiimme ympäristöä kuvaavana säännönmukaisuutena. Mikäli tunnistamme säännönmukaisuuteen liittyvän ärsykkeen uudelleen, emme ole päätöksenteossamme taipuvaisia kognitiiviseen ponnisteluun, vaan toimimme usein kyseenalaistamatta tunnistetun säännönmukaisuuden edellyttämällä tavalla.

Intuition käsite nousee tekoälyn kannalta keskeiseksi koneoppimisen kautta. Murphyn (2014, 1) määritelmä koneoppimiselle löytää datasta automaattisesti säännönmukaisuuksia on yhtenevä Chasen ja Simonin (1973) ihmisen intuition määritelmälle tunnistaa ympäristöstä muistiin varastoituja säännönmukaisuuksia. Sekä tekoälyn koneoppiminen että ihmisen intuitio muotoutuu ympäristöstä eristetyn rajallisen tiedon perusteella. Tekoälyn tapauksessa käytämme tästä tiedosta käsitettä data, kun taas inhimillisen päätöksenteon tapauksessa puhumme yleensä ärsykkeistä. Kuviossa seitsemän havainnollistetaan ympäristön, ympäristöstä rajatun tiedon ja tiedon perusteella muodostettujen säännönmukaisuuksien suhdetta nuolien avulla. Näitä suhteita modifioivat, ihmisen ja tekoälyn päätöksentekoa kuvaavat, kokonaisuudet esitetään kuvion keskellä.

Ympäristöstä valitaan joko harkitusti tai tiedostamatta rajattu määrä tietoa. Kuviossa seitsemän tästä tiedosta käytetään termiä data. Mikäli tiedon luonne mahdollistaa,

muodostuu sen pohjalta säännönmukaisuus, joka ennustaa ympäristön toimintaa. Sekä koneoppimista että ihmisen intuitiota voidaan havainnollistaa kuviossa seitsemän esitettyjen ympäristön, datan ja säännönmukaisuuden suhteina. Vaikka kuvion seitsemän ulkokehällä kuvatun ympäristön, ympäristöstä nostettavan tiedon ja tiedon perusteella muodostettujen säännönmukaisuuksien suhde on ihmisen intuition ja tekoälyn tapauksessa yhteneväinen, eivät näitä suhteita modifioivat tekijät jaa samalla tavalla yhteisiä piirteitä. Mekanismit, joilla tekoäly etsii datassa ilmeneviä säännönmukaisuuksia ja tulkitsee näiden säännönmukaisuuksien perusteella ympäristöään eroavat ihmisen tavasta käsitellä ympäristöstä havaittavia ärsykkeitä ja muodostaa niiden pohjalta säännönmukaisuuksia.

Ihmisen tapauksessa ympäristön, datan ja säännönmukaisuuksien suhteita modifioivat yksilölliset ja organisatoriset vinoumat sekä havaintojen rikastaminen. Nämä ovat kaikki kokonaisuuksia, jotka kuvaavat yksilön tai autonomisesti toimivista yksilöistä koostuvan organisaation suhtautumista ympäristöstä saatavaan tietoon päätöksenteon kontekstissa. Tekoälyn tapauksessa vastaavaa toimijan ja tiedon välistä suhdetta kuvaavat tutkimusraportin liitteissä esitetyt matemaattiset ja tilastolliset mallit. Päätöksenteon viitekehyksessä nämä eivät kuitenkaan pysty juuri tarjoamaan lisäarvoa. Kuten tutkimuksen tekoälyn päätöksentekoa kuvaavassa synteesissä kappaleessa 4.6.4 mainittiin, määrittyy tekoäly päätöksentekijänä, kolmansista osapuolista riippuvaisen luonteensa vuoksi, opetusdatan, opetuksen ja sille annetun tehtävän suhteen. Myös tekoälyn päätöksentekoa kuvaavista ominaisuuksista rajoitettu deskriptiivisyys ja kohdennettu soveltaminen määrittyvät tällöin suhteessa ulkopuoliseen tulkitsijaan. Tämä tekee tekoälyn käsittelyn päätöksenteon näkökulmasta kontekstista irrotettuna entiteettinä mahdottomaksi ja korostaa sovellusympäristön merkitystä tekoälyn päätöksenteon ymmärtämisessä. Kuviossa seitsemän otsikkotasolla mainitut sekä ihmisen että tekoälyn päätöksentekoa ja intuitiota kuvaavat kokonaisuudet on käyty tarkemmin läpi tämän tutkimusraportin toisessa ja neljännessä luvussa.



Kuvio 7: Tekoäly ja ihminen suhteessa ympäristöön, dataan ja säännönmukaisuuksiin

Kun haluamme ymmärtää päätöksentekoa kokonaisuutena tai tarkastella sen onnistumista, emme voi keskittyä pelkästään tarkastelemaan päätöstä tekevän entiteetin ominaisuuksia ja tapaa käsitellä tietoa päätöksenteon yhteydessä. Simon (1990) kuvasi ihmisen rationaalisuutta saksivertauskuvalla, jossa saksien toisen terän muodosti toimijan ajattelu ja kognitiiviset valmiudet ja toisen tehtäväympäristön rakenne. Päätöksentekijä ja ympäristö muodostavat tässä vertauskuvassa erottamattoman kokonaisuuden. Emme pysty ymmärtämään inhimillistä päätöksentekoa pelkästään ihmisille ominaisten vinoumien tai heuristiikkojen kautta. Toisaalta päätöksentekomme ei ole ymmärrettävissä pelkkien ympäristömuuttujien perusteella. Mikäli tahdomme ymmärtää yksilön toimintaa, tulee meidän aina suhteuttaa ihmiselle ominaiset tiedonkäsittelytavat ympäristöön, jossa näitä tapoja sovelletaan. Erityisesti ympäristöriippuvuus on korostuneesti totta myös tekoälyn tapauksessa. Tekoälyn toiminta on selitettävissä vain siinä käyttötarkoituksessa ja ympäristössä, johon järjestelmä on alun perin suunniteltu ja josta järjestelmän opettamiseen käytetty data on kerätty. Tiedon käsittelyn suhteen tekoäly toimii ihmistä orjallisemmin käsittelemällä jokaista yksittäistä datapistettä

identtisesti kohdistamatta erityishuomiota rajattuun määrään datapisteitä. Lisäksi on huomioitavaa, ettei tekoäly aktiivisesti hae ympäristöstään uusia ärsykkeitä vaan toimii ihmisen sille rajaamalla alueella. Reaalimaailman tilanteissa tekoälyn päätöksentekoa ei tällöin voida koskaan ajatella inhimillisestä päätöksenteosta ja sitä kuvaavista kokonaisuuksista irrallisena entiteettinä.

Tekoäly eroaa useista muista teknologioista siinä mielessä, että se muuttaa ihmisen ja teknologian välisen vuorovaikutuksen vaikutusmekanismeja. Teknologiavuorovaikutus on lopputuloksen kannalta merkittävää mitä tahansa teknologiaa tarkasteltaessa. Kuluttaja ei esimerkiksi saa porattua seinään haluamaansa reikää, ellei osaa käyttää hankkimaansa porakonetta. Kuluttajan ja porakoneen välisen vuorovaikutuksen vaikutukset ovat kuitenkin yksisuuntaista. Kuluttajan muutokset tavoissa käyttää teknologiaa eivät vaikuta teknologian tapaan toimia, vaikka niillä on vaikutusta teknologiasta saatavaan hyötyyn. Tekoälyn tapauksessa muutokset teknologian ja ihmisen vuorovaikutuksessa saattavat muuttaa myös teknologian toimintaa. Esimerkiksi tässä tutkimuksessa mainitut Microsoftin chatbot Tay sekä organisaation A tekstiä ymmärtävä ohjelma ovat molemmat sovelluksia, jotka oppivat teknologiavuorovaikutuksen kautta ja muuttavat toimintaansa. Ihmisen ja teknologian vuorovaikutusta ei tekoälyn kannalta voida tarkastella pelkästään teknologiasta saatavan hyödyn maksimoimisen näkökulmasta, vaan sillä on merkittävä vaikutus myös itse teknologian toimivuuteen. Kuviossa seitsemän ihmisen päätöksentekoa kuvaavat yksilölliset ja organisatoriset vinoumat sekä havaintojen rikastaminen on eritelty katkoviivalla tekoälyn päätöksentekoa kuvaavista piirteistä. Tällä pyritään havainnollistamaan niiden toisistaan erottamatonta luonnetta.

Ihmisen tapaan hahmottaa päätöksentekoa edellyttävää ongelmaa vaikuttavat meille ominaiset tavat käsitellä tietoa. Sekä käsiteltävän ongelman rajaus että siitä muodostamamme esiymmärrys syntyvät yksilöllisten ja organisatoristen vinoumien ohjaamina. Kun ongelmaa lähdetään ratkaisemaan, vaikuttaa ymmärryksemme ongelman rajauksesta ja luonteesta siihen, miten kommunikoimme sitä eteenpäin. Ongelman ratkaisun ulkoistamiseksi tekoälyn kanssa käytävää dialogia voidaan verrata kahden ihmisen käymään vastaavaan dialogiin. Samoin kuin toisen ihmisen aiempien kokemusten ja havaintojen kautta muodostuneet yksilölliset vinoumien ilmentymät, vaikuttavat tekoälyn kohdennettu soveltaminen ja rajoitettu deskriptiivisyys välitetyn ja vastaanotetun ongelmasta tehdyn tulkinnan suhteeseen. Tekoälyn päätöksentekoa

kuvaavat kokonaisuudet vuorovaikuttavat inhimillistä päätöksentekoa kuvaavien kokonaisuuksien kanssa ympäristöstä tehtyjen tulkintojen dialogin kautta. Siinä missä ihmisten välisessä dialogissa voimme käyttää kieltä tulkintojen kommunikointiin, tulee meidän tekoälyn tapauksessa turvautua huolelliseen testaamiseen.

Erehtyväisyyden ja päätöksenteon virheellisyyden näkökulmasta teknologiavuorovaikutus nousee keskeiseksi tekijäksi. Tekoäly ei ole ympäristöstään eristetty kokonaisuus, vaan toimii vuorovaikutuksessa ihmisten ja muiden teknologioiden kanssa. Tutkimusraportin viidennessä luvussa esitellyt tekoälyn tekemät virheet nähdään tässä tutkimuksessa syntyvän tekoälyn toiminnan, sen käyttötarkoituksen ja käyttöympäristön sekä teknologiavuorovaikutuksen seurauksena. Kuten Simonin (1990) ihmisen päätöksentekoa kuvaavassa saksivertauskuvassa, kontekstin ymmärtäminen on huomattavan tärkeää myös tekoälyn tapauksessa. Tekoälyä ymmärtääksemme meidän tulee kuitenkin vielä kontekstin ja teknologian lisäksi tarkastella, kuinka teknologia ja ihminen vuorovaikuttavat keskenään juuri kyseessä olevassa kontekstissa.

6.2 Tutkimuksen kontribuutio ihmisen ja koneen erehtyväisyydelle

Tämän tutkimuksen tarkoituksena oli kuvata ja analysoida tekoälyn päätöksentekoa ja erehtyväisyyttä sekä tehdä tekoälyn toiminta tästä näkökulmasta vertailtavaksi ihmisen suoriutumisen kanssa. Tutkimustulokset antavat liikkeenjohdolle lähtökohdan ymmärtää tekoälyä päätöksentekijänä, punnita tekoälyn soveltuvuutta tietyn ongelman ratkaisemiseksi sekä hallita tekoälyn implementointiin liittyviä riskejä. Lisäksi tutkimus auttaa liikkeenjohtoa esittämään sovellusalueen näkökulmasta relevantteja, erityisesti teknologiavuorovaikutusta ja tekoälyn opettamista koskevia, kysymyksiä ja siten arvioimaan ihmisen vaikutusta teknologian käytettävyyteen pitkällä aikavälillä.

Tämän tutkimuksen rooli on pohjustaa yksilöllisemmin ilmiötä tarkastelevaa, tekoälyä päätöksentekijänä käsittelevää myöhempää tutkimusta. Sen lisäksi tutkimuksen on tarkoitus toimia keskustelunavauksena organisaatio- ja tekoälytutkimuksen välillä. Teknologian kehittäjille tutkimus antaa matemaattisen ja ohjelmistotieteellisen lähestymisen rinnalle uuden, käytännön sovellusten kannalta relevantin, viitekehyksen tarkastella teknologian toimintaa. Päätöksenteon ja organisaatiotutkimuksen näkökulmasta tutkimus mahdollistaa tekoälyn käsittelyn inhimilliseen päätöksentekoon verrattavissa olevalla käsitteistöllä. Tällöin sekä inhimillisiä että koneellisia

päätöksentekoentiteettejä sisältävien kokonaisuuksien toiminnan hahmottaminen ja tarkastelu mahdollistuisi helpommin.

Tutkimuksen tulokset auttavat ohjaamaan huomiota teknologiavuorovaikutukseen ja ihmisen ja koneen yhteistoimintaan. Kuten tekoälyn erehtyväisyyttä kuvaavassa luvussa korostetaan, ja toisin kuin inhimillisen erehtyväisyyden tapauksessa, tekoälyn erehtyväisyys ei ole selitettävissä sille ominaisella tavalla käsitellä dataa. Sen sijaan tekoälyn virheet ovat lopputulos käytettävän mallin, suoritettavan tehtävän ja reaalisen toimintaympäristön yhteensovittamisen haasteellisuudesta. Tässä haastavassa kokonaisuudessa teknologiavuorovaikutus näyttäytyy merkittävänä.

6.3 Jatkotutkimusmahdollisuudet

Tämä tutkimus kuvasi ja analysoi tekoälyn päätöksentekoa. Erehtyväisyys ja päätöksenteon virheet nostettiin tutkimuksen rajauksessa korostuneeseen asemaan. Tämä tehtiin osittain tutkijan oman mielenkiinnon ohjaamana mutta osin myös siksi, että ne muodostivat inhimillistä päätöksentekoa kuvaavassa kirjallisuudessa ja tutkimuksessa merkittävän kokonaisuuden. Tutkimus lähestyi tekoälyn tekemiä virheitä tapausluontoisesti esittelemällä ja analysoimalla julkisuudessa tunnettuja tilanteita, joissa tekoäly oli erehtynyt. Tapausluontoisen virheiden analysoimisen ohella jatkotutkimuksen olisi hyvä ottaa tekoälyn erehtyväisyyteen kantaa myös kvantitatiivisesti. Erilaisten virhetyyppien tilastollinen taulukoiminen on edellytys, mikäli haluamme luoda tietoa tekoälyn erehtyväisyyden kustannuksista, arvioida teknologiahankkeiden taloudellista riskiä ja identifioida ympäristöjä, joissa teknologian implementointi on osoittautunut haasteellisimmaksi.

Erehtyväisyyden lisäksi jatkotutkimuksen olisi aiheellista lähestyä tekoälyn päätöksentekoa myös muista lähtökohdista. Esimerkiksi tekoälyn päätöksenteon tehokkuuden tai tarkkuuden tarkastelulla saavutettaisiin arvokasta lisätietoa teknologian sovellettavuuspotentiaalin arviointiin. Erehtyväisyydelle vaihtoehtoisia lähestymistapoja olisi hyvä tarkastella sekä laadullisesti että määrällisesti ja sekä yleisellä tasolla että kapeammin rajatuilla alueilla.

Teknologiaan toimintaan ja suoriutumiseen keskittyvän tutkimuksen lisäksi tämän tutkimuksen löydösten perusteella on aiheellista rakentaa syvempää ymmärrystä teknologian toiminnasta ja käyttäytymisestä sovelluskontekstissa. Erityisesti

teknologiavuorovaikutuksen tarkempi tutkiminen tekoälyn toiminnan näkökulmasta on merkittävä jatkotutkimuskokonaisuus tekoälyn kolmansista osapuolista riippuvaisen luonteen vuoksi. Teknologisoituminen, osana ihmisistä koostuvia kokonaisuuksia, lisää myös päätösten automatisointiin liittyvien eettisten kysymysten tarkastelun merkitystä.

Tässä tutkimuksessa kuvattiin tekoälyä päätöksenteon näkökulmasta yleistasolla. Tutkimuksen löydökset korostavat tehtävän, käyttöympäristön ja teknologiavuorovaikutuksen roolia osana tekoälyn päätöksenteon ymmärtämistä. Nämä löydökset auttavat ohjaamaan tulevan tutkimuksen fokusta, mutta ne eivät vielä anna teknologian kehittäjille tai sen implementoijille tarpeeksi yksityiskohtaista tietoa tekoälyn soveltamisesta juuri halutulla alueella. Tämän tutkimuksen laajan ja yleisluontoisen kuvauksen tarkentaminen kapeammille ja spesifimmeille sovellusalueille kuvaa tämän tutkimuksen kaikkia potentiaalisia jatkotutkimussuuntia.

LÄHTEET

Kirjallisuuslähteet

Arkes, H., R. ja Blumer, C. (1985). The Psychology of Sunk Cost. *Organizational Behavior and Human Decision Processes*, 35, 124 – 140.

Antonakis, J. ja Day, D. (2012). *The Nature of Leadership.* Los Angeles: SAGE Publications

Church, A. (1936). An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics*, 58, 345 – 363.

Manning, C.D. ja Schütze, H., (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Bateman (1983). Resource allocation after success and failure: the role of attributions of powerful others and probabilities of future success. *Working Paper.* Department of Management, Texas A&M University.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, 3, 439–449.

Boole, G (1854). *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. Dover Publications.

Bowen, M. G. (1987). The escalation phenomenon reconsidered. Decision dilemmas or decision errors? *Academy of Management Review*, 12, 52–66.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

Brighton, H. (2006). Robust Inference with Simple Cognitive Models, Teoksessa C.Labiere ja B. Wray (toim.), *Between a rock and a hard place: Cognitive science principles meet AI-hard problems*. Papers from AAAI Spring Symposium

Brockner, J., & Rubin, J. Z. (1985). *Entrapment in escalating conflicts: A social psychological analysis*. New York: Springer-Verlag.

Brockner, Rubin ja Lang (1981). Face-saving and entrapment, *Journal of Experimental Social Psychology* 17, 68–79.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, 90, 63–86.

Christensen-Szalanski, J. (1978). Problem solving strategies: A selection mechanism, some implications, and some data. *Organizational Behavior & Human Performance*, 22(2), 307–323.

Chowdhury, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37, 51–89.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess, Teoksessa W. G. Chase (toim.) *Visual information processing*. New York: Academic Press, 215–281.

Cyert, R.M., ja March, J.G. (1959). A Behavioral Theory of Organizational Objectives. Teoksessa *Modern Organizational Theory,* New York: Wiley, 76-90.

Cyert, R.M., ja March, J.G. (1963). *A Behavioral Theory of the Firm.* Englewood Cliffs: Prentice-Hall

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? Teoksessa *Simple heuristics that make us smart.* Oxford: Oxford University Press, 97–118.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582.

Davis, M. (2000). *The Universal Computer: The Road from Leibniz to Turing*, New York: W. W. Norton & Co.

DeMiguel, V., Garlappi, L., Uppal, R. (2009). Optimal Versus Naïve Diversification: How Inefficient is the 1/N Portfolio Strategy. *The Review of Financial Studies* 22(5).

DeGroot, A. D. (1978). *Thought and choice in chess.* Haag: Mouton.

Drummond, H., (1994). Too Litle Too Late: A Case Study of Escalation in Decision Making, *Organization Studies,* 15(4), 591–607.

Dubois, A. ja Gadde, L-E. (2002). Systematic Combining: An Abductive Approach to Case Research. *Journal of Business Research*, 55(7), 553–560.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Eriksson, P. ja Kovalainen, A. (2008). *Qualitative Methods in Business Research*. SAGE Publications.

Eskola, J. ja Suoranta, J. (1998). *Johdatus laadulliseen tutkimukseen*. Vastapaino

Fausett, L (1994). Fundamentals of Neural Networks. New Jersey: Prentice Hall.

Fox, F., ja Staw B. (1979). The Trapped Administrator: Effects of Job Insecurity and Policy Resistance upon Commitment to a Course of Action. *Administrative Science Quarterly,* 24(3), 449–471.

Ge, Z.M. ja Lee C.I. (2005). Anticontrol and synchronization of chaos for an autonomous rotational machine system with a hexagonal centrifugal governor. *Journal of Sound and Vibration,* 282, 635–648.

Gigerenzer, G. (2008). Why Heuristics Work, *Perspectives on Psychological Science* 3, 20–29.

Gigerenzer, G. ja Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science* 1, 107 – 143.

Gigerenzer, G. ja Goldstein, D. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality, *Psychological Review* 103, 650 – 669.

Goldstein, D., Gigerenzer, G. (2002), Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review*, 109, 75–90.

Harari, Y. (2015). Sapiens: A Brief History of Humankind. Harper.

Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. Macmillian College Publishing Company, Inc, 1994.

Hirsjärvi, S. ja Hurme, H. (2011). *Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö*, Gaudeamus

Hirsjärvi, S., Remes, P., Sajavaara, P. (2009). *Tutki ja kirjoita.* Helsinki: Tammi.

Hogart R., Karelaia N., (2007). Heuristic and Linear Models of Judgement: Matching Rules and Environments. *Psychological Review* 114(3), 733–758.

Hyvärinen, M., Nikander, P. & Ruusuvuori, J. (2017). *Tutkimushaastattelun käsikirja.* Tampere: Vastapaino.

Jacoby, L., Dallas, M. (1981). On the Relationship Between Autobiograpical Memory and Perceptual Learning. *Journal of Experimental Psychology*,110(3), 304–340.

Jang J., Sun, C., Mizutani, E. (1997). *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall.

Johnson, J. ja Goldstein, D. (2004). Defaults and Donation Decisions. *Transplantation,* 78(12), 1713–1716.

Kahneman, D. (2003). Autobiography. *Les Prix Nobel 2002* Stockholm: Almqvist & Wiksell International.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, D. ja Klein, K. (2009). Conditions for Intuitive Expertise. *American Psychological Association*, 64(6), 515–526.

Kahneman, D. ja Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica,* 47, 263–292.
Kahneman, D. ja Tversky, A. (1981). The Framing of Decisions and the Psychology of Choice. *Science,* 211, 453–458.
Kahneman, D. ja Tversky, A. (1982). The psychology of preferences. *Scientific American* 246, 160-73.
Kahneman, D. ja Tversky, A. (1984). Choices, values, and frames. *American Psychologist,* 39, 341–350.
Laine, T. (2018). Miten kokemusta voidaan tutkia?: fenomenologinen näkökulma. Teoksessa R. Valli (Toim.), *Ikkunoita tutkimusmetodeihin 2. Näkökulmia aloittelevalle tutkijalle tutkimuksen teoreettisiin lähtökohtiin ja analyysimenetelmiin*. Jyväskylä, Finland: PS-kustannus, 29–50.
Louridas, P. ja Ebert, C. (2016). Machine Learning, *IEEE Software*, September/October, 110–115.
Kirsch. R. A. (1954). Experiments with a Computer Learning Routine. *Computer Seminar Notes*, Heinäkuu 30.
Koskinen, I., Alasuutari, P., Peltonen, T. (2005). *Laadulliset menetelmät kauppatieteissä*, Vastapaino.
Maimon, O. & Rokach, L. (2008). *Data Mining with Decision Trees: Theory and Applications.* World Scientific Publishing.
March, J.G. (1962). Some Observations on Political Theory. Teoksessa L.K. Caldwell (toim.) *New Viewpoints on Politics and Public Affairs.* Bloomington: IN:University of Indiana Press, 121–139.
March, J.G. (1981). Footnotes to Organizations and Theories of Choice. *Administrative Science Quarterly,* 26(4), 563–77.
March, J.G. (1988). *Decisions and Organizations*. Oxford: Wiley-Blackwell
March, J.G, Lee, S., Sproull, M. (1991). Learning from Samples of One or Fewer. *Organization Science,* 2(1), 1–13.
March,J.G., ja Simon, H.A. (1958). *Organizations.* New York, NY:Wiley.
March, J.G., ja Olsen, J.P. (1975). The Uncertainty of the Past: Organizational Learning Under Ambiguity. *European Journal of Political Research,* 3(2), 147–171.
March, J.G. ja Shapira, Z. (1987). Managerial Perspectives on Risk and Risk Taking. *Management Science,* 33(11),1404–1418.
Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.
Miles, M., Huberman, M. (1994). *Qualitative Data Analysis*. Sage Publications.
Minsky, M. (1968). *Semantic Information Processing*. Cambridge: MIT Press.
Murphy, K. (2014). *Machine Learning: A Probabilistic Perspective*, MIT Press
Nils, J. Nilsson (2010). *The Quest for Artificial Intelligence A History of Ideas and Achievements*. Cambridge: University Press.
Northcraft, G.B. ja Wolf, G. (1984). Dollars, sense and sunk costs. A life cycle model of resource allocation decisions. *Academy of Management Review,* 9(2), 225–234.
Oxford (2017). Artificial intelligent. *A Dictionary of Computer science*. Oxford University Press. saatavissa: http://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975-e-204?rskey=ClQBQw&result=202, viitattu 19.9.2017.
Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology*, 135, 207–236.

Saarela-Kinnunen, M & Eskola, J. (2010). Tapaus ja tutkimus = tapaustutkimus? teoksessa: Aaltola & Valli (toim.) *Ikkunoita tutkimusmetodeihin 1*. Jyväskylä: PSkustannus, 189–199.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, 252–266.

Shanteau, J., Thomas, R.P., (2000). Fast and frugal heuristics: What about unfriendly environments?, *Behavioral and Brain Sciences* 23, 762–763.

Simon H.A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics,* 69(1), 99–118.

Simon, H.A., (1956). Rational Choice and the Structure of the Environment. *Psychological Review* 63, 129–138.

Simon, H. (1957). *A Behavioral Model of Rational Choice, Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.

Simon, H.A., (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1–19.

Staw, B. (1976). Knee-Deep in the Big Muddy: A Study of Escalating Commitment, *Organizational Behavior and Human Performance,* 16, 27–44.

Staw, B. (1980). Rationality and justification in organizational life. Teoksessa Research in organizational behavior. Greenwich: JAI Press.

Staw, B. (1981). The Escalation of Commitment To a Course of Action. *Academy of Management Review*, 6(4), 577 – 587.

Staw, B. ja Fox, F. (1977). Escalation: The Determinants of Commitment to a Chosen Course of Action, *Human Relations,* 30, 431–450.

Staw, B. & Ross, J. (1978). Commitment to a policy decision: A multi-theoretical perspective. *Administrative Science Quarterly,* 23, 40-64.

Staw, B. M., & Ross, J. (1980). Commitment in an experimenting society: A study of the attribution of leadership from administrative scenarios. *Journal of Applied Psychology,* 65(3), 249–260.

Staw, B. M., & Ross, J. (1987). Behavior in escalation situations: Antecedents, prototypes, and solutions. *Research in Organizational Behavior*, 9, 39-78.

McCulloch, W., and Walter Pitts (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.

Turing, A.M., (1936). *On Computable Numbers, with an Application to the Entscheidungsproblem*. Proceedings of the London Mathematical Society, Series 2, 42, 230–265.

Turing, A.M. (1946). Seminaariteksti, Alan Turingin ACE raportti 112, New York: Basic Books.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124−1131.

Tversky, A. ja Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59 (4) 251–278.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker. New York: Cambridge University Press.

Pyle, D. (1999). *Data Preparation for Data Mining*. Morgan Kaufmann Publishers.

Silverman, D. (2010). *Doing qualitative research: A practical handbook*. Sage Publications.

Teger, A. I. (1980). *Too Much Invested to Quit*, Elsevier.

Torrey ja Shavlik (2009). Transfer Learning.Teoksessa *Handbook of Research on Machine Learning Applications,* Serrano: IGI Global.
Tuomi, J. ja Sarajärvi, A. (2004). Laadullinen tutkimus ja sisällönanalyysi. Jyväskylä: Gummerus Kirjapaino Oy.
Saaranen-Kauppinen, A. ja Puusniekka, A. (2006). Menetelmäopetuksen Tietovaranto KvaliMOTV. Tampere: Yhteiskuntatieteellinen tietoarkisto
Rowley, J. (2012). Conducting research interviews. *Management Research Review*, 35(3), 260 –271.
Whyte, G. (1986). Escalating Commitment to a Course of Action: A Reinterpretation. *The Academy of Management Review*, 11.
Whyte, G. (1990). Entrapment: Are groups less susceptible? Seminaaripaperi: the Academy of Management. San Francisco.
Yin, R. K., (2009). Case study research: Design and methods. Thousand Oaks: SAGE.

Verkkolähteet

Antonaci Z., (2016). NPC AI update, Frontierin keskustelufoorumi. saatavissa: https://forums.frontier.co.uk/showthread.php?t=258662.viitattu: 21.10.2018.
Clark, B. (2017). Facebook's AI accidentally created its own language. saatavissa: https://thenextweb.com/artificial-intelligence/2017/06/19/facebooks-ai-accidentally-created-its-own-language/#.tnw_mWKGOzpB. viitattu: 10.9.2018
Eremenko (2018). Machine Learnin A-Z: Hands-On Python & R In Data Science. verkkokurssi. saatavissa: https://www.udemy.com/machinelearning/, viitattu: 10.9.2018.
Dieterich, W., Mendoza, C., Brennan, T. (2016). COMPAS Risk Scales Demonstrating Accuracy Equity and Predictive Parity. Northpointe. saatavissa: https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html. viitattu: 20.10.2018.
Gartner (2016). Gartner Hype Cycle for Emerging Technologies, 2016. saatavissa: https://www.gartner.com/smarterwithgartner/3-trends-appear-in-the-gartner-hype-cycle-for-emerging-technologies-2016/. viitattu: 20.9.2018
Gartner (2017). Gartner Hype Cycle for Emerging Technologies, 2017. Saatavissa: https://www.forbes.com/sites/gartnergroup/2017/08/18/future-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/#23c6c9b74b97. viitattu: 20.9.2018
Gartner (2018). Gartner Hype Cycle for Emerging Technologies, 2018. saatavissa: https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/. viitattu: 20.9.2018
Kirkpatrick, K. (2016). Battling Algorithmic Bias. *Communications of the ACM.* saatavissa; https://cacm.acm.org/magazines/2016/10/207759-battling-algorithmic-bias/abstract. viitattu: 30.10.2018.
Levis, M., Yarats, D., Dauphin, Y., Parikh, D., Batra, D., (2017). Deal or No Deal? End-to-End Learning for Negotiation Dialogues. saatavissa: https://arxiv.org/pdf/1706.05125.pdf. viitattu: 30.10.2018.
Nieva, R., (2017). Facebook put cork in chatbots that created a secret language. *cnet.* saatavissa: https://www.cnet.com/news/what-happens-when-ai-bots-invent-their-own-language/. viitattu: 30.10.2018
Price, R., (2016). Microsoft Took Its New A.I. chatbot Offline After It Started Spewing Racist Tweets, Slate. saatavissa: https://slate.com/business/2016/03/microsoft-s-new-ai-chatbot-tay-removed-from-twitter-due-to-racist-tweets.html. viitattu: 23.10.2018.

ProPublica (2016a). Machine Bias. saatavilla: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. viitattu: 20.10.2018.
ProPublica (2016b). Technical Response to Northpointe. saatavilla: https://www.propublica.org/article/technical-response-to-northpointe. viitattu: 20.10.2018.
Reese (2016). Tesla driver dies in first fatality with Autopilot: What it means for the future of driverless cars, TechRepublic. saatavissa: https://www.techrepublic.com/article/tesla-driver-dies-in-first-fatality-with-autopilot-what-it-means-for-the-future-of-driverless-cars/. viitattu: 21.10.2018.
The Tesla Team (2016). A Tragic Loss. Teslan blogi postaus. saatavissa: https://www.tesla.com/blog/tragic-loss. viitattu: 21.10.2018.
Thompson (2016). Here's how Tesla's Autopilot works, businessinsider. saatavissa: https://www.businessinsider.com/how-teslas-autopilot-works-2016-7?r=US&IR=T. viitattu: 21.10.2018.
Tuomivaara, T. (2005). Tieteellisen tutkimuksen perusteet, kurssimateriaali. saatavissa: https://docplayer.fi/22096431-6-kvantitatiivinen-ja-kvalitatiivinen-tut-kimus.html, viitattu: 18.10.2018.
Vincent, J., (2016a). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. saatavilla: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist. viitattu: 22.10.2018.
Vincent, J., (2016b). Facebook blocks insurer exploiting user data to find "conscientieous" drivers. *The Verge*. saatavissa: https://www.theverge.com/2016/11/2/13496316/facebook-blocks-car-insurer-from-using-user-data-to-set-insurance-rate. viitattu: 24.10.2018.
Yin-Poole, W. (2016). Elite: Dangerous' latest expansion caused AI spaceships to unintentionally create super weapons. *Eurogamer*. saatavissa: https://www.eurogamer.net/articles/2016-06-03-elite-dangerous-latest-expansion-caused-ai-spaceships-to-unintentionally-create-super-weapons. viitattu 21.10.2018.

Henkilölähteet

Huttunen, H. (2018). *Haastattelu*, 31.5.2018
Lehti, R. (2018). *Haastattelu*, 14.11.2018
Nimetön asiantuntija. *Haastattelu*, 15.11.2018
Nimetön asiantuntija. *Haastattelu*, 15.11.2018
Nimetön asiantuntija. *Haastattelu*, 15.11.2018

Liite 1. Tekoälytutkimuksen pohja

Tekoälytutkimuksen eriytymistä omaksi tutkimusalueekseen on edeltänyt edistysaskeleet useilla eri tieteenaloilla. Tässä liitteessä esitellään nykymuotoiselle tekoälytutkimukselle keskeisimmät askeleet logiikan, tilastotieteen, psykologian, neurotieteiden ja insinööritieteiden alueilla.

3.1.1 Logiikka ja tilastotiede

Ensimmäiset viitteet päättelyn ja älykkyyden automatisointiin saatiin noin 300eaa. kun kreikkalainen filosofi Aristoteles pyrki analysoimaan ja kodifioimaan loogista päättelyä. Aristoteles identifioi loogisen syllogismin, jossa tiettyjen tosi premissien olemassaolo, johtaa itsessään päätelmään. (Nilsson 2010, 27) Nilsson (2010, 27) antaa esimerkin eräästä tunnetusta syllogismista:

1) *Kaikki ihmiset ovat kuolevaisia (tosi premissi)*
2) *Kaikki kreikkalaiset ovat ihmisiä (tosi premissi)*
3) *Kaikki kreikkalaiset ovat kuolevaisia (päätelmä)*

Juuri syllogismin muoto tekee siitä tekoälytutkimuksen kannalta merkittävän. Logiikkaketju ei ole tilannesidonnainen, vaan kuolevaisuus, ihmiset ja kreikkalaiset voidaan korvata universaaleilla symboleilla:

1) $B \in A$ *(tosi premissi)*
2) $C \in B$ *(tosi premissi)*
3) $C \in A$ *(päätelmä)*

Nilssonin (2010, 28) mukaan kaksi tekijää syllogismeissa antaa viitteitä päättelyn automatisoinnin mahdollisuuteen: väittämät voidaan esittää taulukoidussa muodossa symbolein (1) ja päätelmät voidaan muodostaa premisseistä kohdistamalla symboleihin tiettyjä operaatioita (2). Universaalien symbolien, sekä erilaisten operaatioiden käyttö muodostaa myös monen modernin tekoälysovelluksen ytimen. Logiikan tutkimus on sittemmin ottanut suurimmat harppauksensa 1600-luvun ja 1800-luvun puolenvälin välillä, jolloin erilaiset yhtälömuotoiset operaatiot kehittyivät (Nilsson, 2010, 28).

Aristoteleen jälkeen Wilhelm Leibniz (1646 – 1716) oli eräs ensimmäisistä loogisen päättelyn parissa työskentelevistä tieteilijöistä. Leibnizin pyrkimyksenä oli päättelyn koneellistaminen. Hän pyrki kehittämään universaalia merkistöä, jonka avulla

pystyttäisiin esittämään kaikki tietoisuus. Leibniz uskoi, että aivan kuten sanat voitiin koostaa kirjaimista, myös ajatukset voitaisiin muodostaa omaa merkistöään käyttäen. Mikäli merkistö kyettäisiin lisäksi esittämään numeroin, voitaisiin päätelmät tarkistaa yksinkertaisesti ”laskemalla”. Leibnizin ajatuksen pääongelmaksi muodostui tällaisen universaalin merkistön keksiminen. Siitä huolimatta hänen työnsä logiikan parissa antoi tärkeitä vihjeitä päättelyn mekanisoimiseksi. (Davis 2000)

1854 George Boole julkaisi kirjan *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities.* Useiden muiden ajan tieteilijöiden tapaan Boole pyrki esittämään inhimillisen päättelyn erityispiirteitä ja esittämään ne matemaattisessa muodossa. Esimerkiksi hänen neljäs ajattelua kuvaava premissi koski vastakohtaisuutta ja määritteli, ettei mikään olevainen voinut omistaa tiettyä ominaisuutta samalla ollen omistamatta sitä. Algebrallisessa muodossa Boole esitti premissin seuraavasti

$$x(1-x)=0$$

Missä x edustaa mitä tahansa luokkaa objekteja, $(1-x)$ luokan vastaluokkaa ja 0 luokkaa, jota ei ole olemassa. (Boole 1854)

Boolen työn seurauksena syntyneessä Boolean algebrassa 0 edustaa *epätotta* ja 1 *totta*. Boolean algebran perus-operaatiot ovat TAI ja JA, joita merkitään merkeillä $+$ ja $\times$. Operaattoreiden laskusääntöjä hyödyntäen voidaan laskea eri väittämien (esim. q ja p) totuusarvoja. Perus operaatioiden laskusäännöt tunnetaan seuraavina

$$1+0=1$$

$$1\times 0=0$$

$$1+1=1$$

$$1\times 1=1$$

$$0+0=0$$

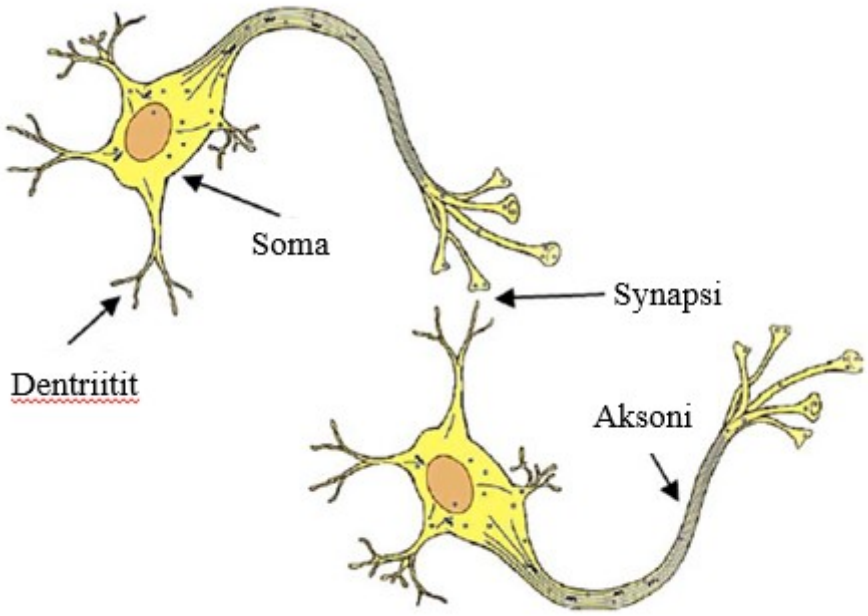
ja

$$0\times 0=0$$

Boolean algebra näytteli myöhemmin tärkeää roolia puhelimen ja tietokoneen kehittämisessä. Ennen kaikkea Boolen työ näytti, että loogisia operaattoreita käyttäen kyettiin rajalliseen loogiseen päättelyyn. (Nilsson 2010, 28-32)

3.1.2 Neurotieteet ja psykologia

1800-luvun lopulla ja 1900-luvun alussa neurotutkijat olivat löytäneet aivojen toiminnan kannalta merkittävän neuroniksi nimetyn solun (Nilsson 2010, 34). Neuronin rakenne koostuu soomasta, dentriiteistä sekä aksonista. Sooma on hermosolun runko-osa ja dentriittien ja aksonin risteymakohta, joka vastaa suurimmasta osasta solun toiminnoista. Dentriitit ovat solun tuojahaarakkeita, jotka toimivat solun pääasiallisena tiedonkeräämisvälineenä. Aksoni eli solun viejähaarake kuljettaa soluun tulleen hermoimpulssin eteenpäin hermoliitoksen (synapsi) välityksellä. Hermosolun rakenne on estetty kuviossa kahdeksan



Kuvio 8: Neurosolun rakenne (Nilsson 2010, 36)

Vuonna 1943 Neuropsykologi Warren McCulloch ja matemaatikko (logician) Walter Pitts esittivät yksinkertaisen matemaattisen mallin neuronin toiminnasta ja osoittivat, että verkko tällaisia malleja kykeni suoriutumaan kaikista mahdollisista laskennallisista operaatioista. McCulloch-Pitts-neuroni oli matemaattinen input-output-abstraktio, jossa jokainen vaste sai joko arvon 1 tai 0. Verkko muodostettiin siten, että jokaisen McCulloch-Pitts-neuronin vaste toimii seuraavan neuronin syötteenä. Mikäli kaikkien

syötteiden yhteenlaskettu arvo yrittää määritellyn tason, McCulloch-Pitts-neuroni aktivoituu. (McCulloch ja Pitts 1943)

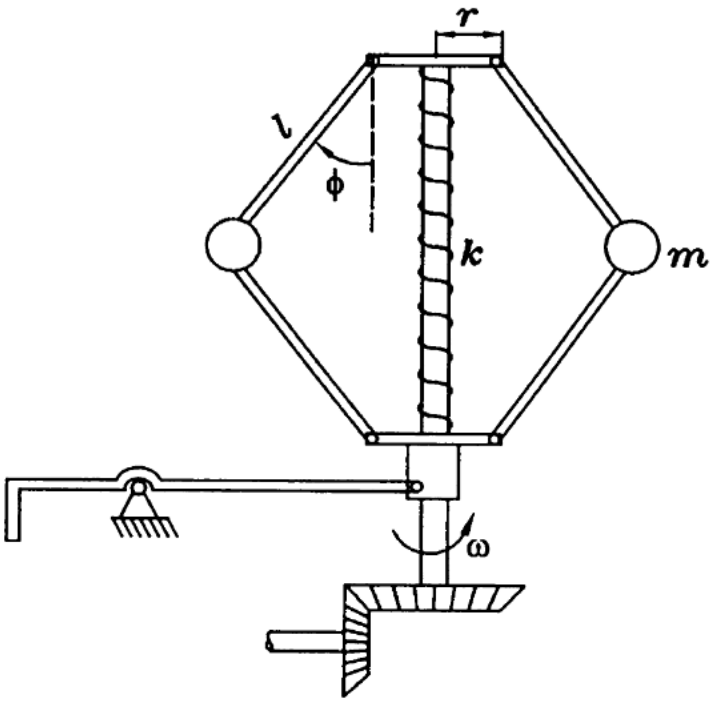
1900-luvun puolivälissä psykologian tutkimuksessa alkoi nousta uusi behavioristinen suuntaus, jonka mukaan psykologian tulisi olla enemmän käyttäytymisen, kuin mielen tutkimusta (Antonakis ja Day 2012). Behavioristinen suuntaus ei esimerkiksi pitänyt uskomusten, halujen tai tavoitteiden identifioimisen korostamista tutkimuksessa tärkeänä (Nilsson 2010, 37). Tekoälytutkimuksen näkökulmasta merkittävimpiä suuntauksen edustajia oli B.F Skinner. Skinnerin tavoitteena oli viedä psykologiaa luonnontieteellisempään suuntaan keskittymällä puhtaasti ilmiöihin, jotka olivat mitattavissa, kuten tiettyä ärsykettä seuraaviin reaktioihin. (Nilsson 2010, 37) Skinnerin työtä ja behavioristista, suuntausta on myöhemmin kritisoitu muun muassa siitä, etteivät vahvan tilastolliset teoriat pystyneet juurikaan selittämään kuvaamiaan ilmiöitä (Minsky 1968, 2).

Behavioristisen suuntauksen ja erityisesti B.F. Skinnerin (1904-1990) merkittävin perintö tekoälytutkimukselle oli vahvistusoppiminen, joka perustuu ajatukseen stimulaatiovasteen tukemisesta. Mikäli ympäristössä toimivan entiteetin vastetta tuetaan palkitsemalla, nostaa se kyseisen vasteen todennäköisyyttä vastaavissa olosuhteissa myös tulevaisuudessa. Vahvistettu oppiminen nousikin suosituksi strategiaksi tekoälytutkijoiden keskuudessa. (Nilsson 2010, 40) Vuoden 1954 seminaarimuistiossa Russell Kirsch kuvaili kuinka hänen luomansa ”keinotekoinen eläin” kykeni käyttämään vahvistettua oppimista oppiakseen oikean liikkeen pelissä. ”Eläin” mallinsi stimulaatiota seuranneen vastapuolen liikkeen. Kun sama stimulaatio seuraavan kerran tapahtui, toisti eläin vastustajansa samassa tilanteessa tekemän liikkeen. Mitä useammin vastustaja toisti samaa liikettä tietyssä tilanteessa, sitä paremmin ”eläin” ehdollistui kyseiseen liikkeeseen. (Kirsch 1954)

3.1.3 Insinööritieteet ja tietokoneen kehitys

Erilaisia toimintoja automatisoivia laitteita on ollut olemassa jo pitkään. Eräitä varhaisimmista esimerkeistä ovat myöhäisellä keskiajalla italialaisten kaupunkien torneihin rakennetut kellot. Kellot tarvitsivat ulkopuolisen energianlähteen, kuten painoja, jousia tai ihmisiä, mutta pystyivät muuten toimimaan täysin automaattisesti. Ensimmäisten automatisoitujen järjestelmien keskeisin rajoite oli, etteivät ne kyenneet minkäänlaiseen vuorovaikutukseen ympäristönsä kanssa. (Nilsson 2010, 46)

Ympäristöön reagoimisen edellytyksenä voidaan pitää palaute-kontrolli-funktiota (*feedback control),* joka antaa jonkin koneen toiminnan elementin, kuten käyntinopeuden syötteenä takaisin koneelle. (Nilsson 2010, 46-47) Nilsson (2010, 48-49) käyttää James Wattin vuonna 1788 kehittämää, höyrykoneen käyntinopeuden tasaavaa, keskipako-ohjainta (kuvio 9) esimerkkinä palaute-kontrolli-funktiosta. Keskipako-ohjaimen akselin k kulmanopeuden ω kasvattaminen saa m massaiset pallot etääntymään pyörimisakselista. Kulmanopeuden pienentäminen taas saa pallot lähenemään akselia. Kulmanopeuden muutoksesta johtuva pallojen liike aiheuttaa kulman ϕ muutoksen. (Ge ja Li 2005, 636) Wattin höyrykonetta ohjaavassa mekanismissa akseli k on kytketty höyrykoneeseen siten, että sen kulmanopeus määräytyy höyrykoneen käyntinopeuden mukaan. Kulman ϕ muutos ohjaa venttiiliä, joka säätelee koneeseen virtaavan ilman määrää. Käyntinopeuden nouseminen nostaa akselin k kulmanopeutta. Tämä saa pallot etääntymään akselista, mikä pienentää höyrykoneeseen virtaavan ilman määrää ja laskee käyntinopeutta. Käyntinopeuden laskiessa pallot lähenevät akselia, mikä taas johtaa syöttöilman lisäämiseen ja käyntinopeuden nousuun. (Nilsson 2010, 49) Wattin keskipako-ohjain on palaute-kontrolli-funktio, joka käyttää koneen omaa käyntinopeutta syötteenä koneen ohjaamiseen.



Kuvio 9: Keskipako-ohjain (Ge ja Lee 2005, 636)

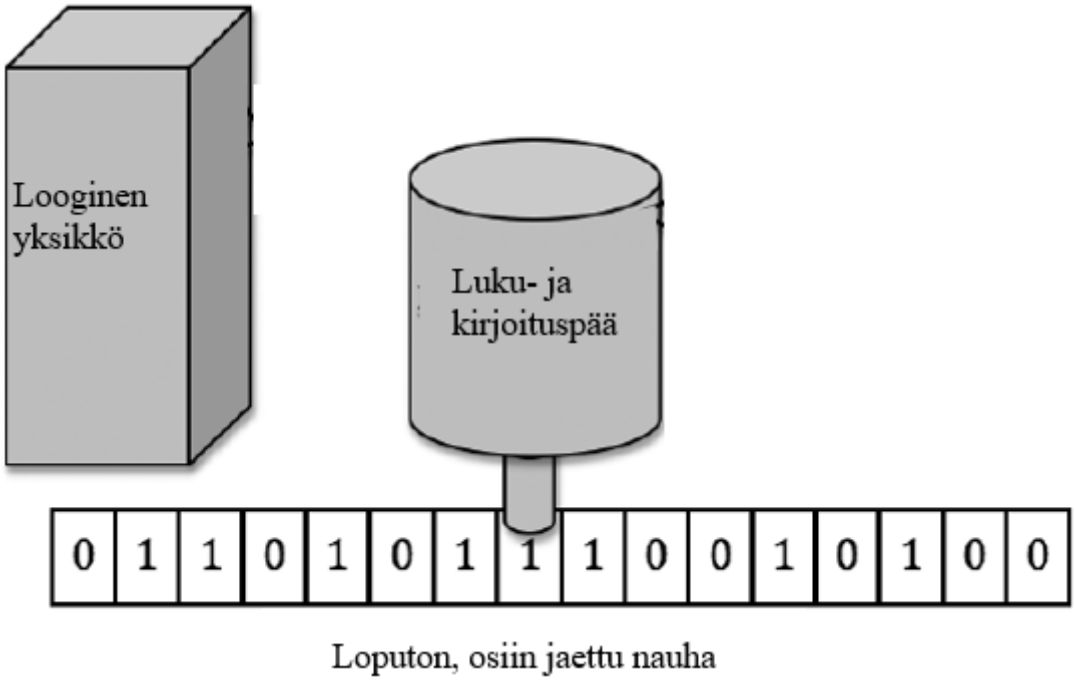
1943 Norbert Wiener ja Julian Bigelow keksivät termin kybernetiikka kuvaamaan palautteeseen ja kontrolliin liittyviä kokonaisuuksia. Sana juontuu kreikan kielen sanasta *kybernetike,* joka tarkoittaa ohjaamisen taitoa. Vaikka myöhemmin etuliitettä kyber on käytetty hyvin laajasti kuvaamaan tietokoneisiin, robotiikkaan ja internettiin liittyviä ilmiöitä, tarkoittaa termi teknisesti edelleenkin järjestelmien palautteeseen ja ohjaamiseen liittyviä toimintoja.

Muun muassa Leibnizin ja Boolen ja tilastotieteellisiä löydöksiä voidaan pitää aikaisina yrityksinä rakentaa pohjaa älykkäälle ohjelmistolle. Päättely ja älykäs toiminta kuitenkin edellyttävät ohjelmiston lisäksi fyysistä konetta. Ihmisillä ja eläimillä kyseinen kone on aivot. Tämän päivän tekoälysovelluksille vastaavana fyysisenä elementtinä toimii tietokone. (Nilsson 2010, 53)

Laskentaan suunnitelluilla laitteilla on pitkä historia. Ensimmäiset laskentaan käytettävät apuvälineet olivat täysin mekaanisia. Niitä ei voinut ohjelmoida ja ne kykenivät vain tiettyyn ennalta määriteltyyn operaatioon kuten kerto-, jako-, yhteen- tai vähennyslaskuun. Ensimmäinen ohjelmoitava tietokone suunniteltiin 1834-1837. Messinkirattaiden ohjaamiseen perustuvaa, höyryvoimaa energianlähteenä käyttävää tietokonetta ei kuitenkaan koskaan rakennettu projektin kaaduttua rahoitusongelmiin. Ohjelmoitavan yleistietokoneen kehitys joutuikin odottamaan sähkön keksimistä. Ensimmäiset ohjelmoitavat useampiin operaatioihin kykenevät yleistietokoneet toteutettiin sähkö-mekaanisilla releillä, jotka kuitenkin pian korvattiin elektroniputkilla, sillä ne takasivat nopeamman ja luotettavamman laskemisen. Tämän päivän tietokoneissa elektroniputket ovat korvautuneet miljardeilla piirilevyyn asennetuilla transistoreilla. (Nilsson 2010, 53-56)

Ennen varsinaisen tietokoneen rakentamista käytiin keskustelua ongelmista, joita kyettäisiin käsittelemään laskennallisesti. Vuonna 1936 Alonzo Church löysi laskettavissa olevien funktioiden joukon, joista hän käytti nimeä *recursive* (Church 1936). Samoihin aikoihin Alan Turing kuvaili kuvitteellista, laskentaan kykenevää konetta. Myöhemmin Turingin koneena tunnettu kone koostui kolmesta komponentista: (1) Loputon, osiin jaettu nauha, jonka kuhunkin osaan on tallennettu joko arvo 1 tai 0. (2) Luku ja kirjoituspää, joka lukee nauhan arvoja ja kykenee kirjoittamaan niiden päälle. (3) Looginen yksikkö, joka voi muuttaa koneen statusta nauhalta luetun arvon perusteella,

ohjata kirjoituspäätä, siirtää nauhaa yhden pykälän eteen tai taakse ja lopettaa operaation suorittamisen. (Turing 1937) Turingin kone on kuvattu kuviossa 10.



Kuvio 10: Turingin kone

Turing esitti myös ajatuksen myöhemmin ekvivalenssiluokkana tunnetun joukon olemassaolosta. Ekvivalenssiluokalla Turing kuvasi kuvitteellisella koneellaan laskettavissa olevien numeroiden joukkoa ja esitti sen olevan yhteneväinen Churchin rekursiivi-joukkoon. (Turing 1936) Tätä Church-Turing teesiksi nimettyä lakia ei olla kyetty todistamaan, mutta sitä pidetään yleisesti pitävänä, eikä vielä ole kyetty löytämään esimerkkitapausta, joka rikkoisi säännön. (Nilsson 2010, 56) Ekvivalenssiluokan esittelemisen lisäksi Turing todisti, että Turingin koneen logiikkayksikkö kyettiin määrittelemään jokaiselle laskettavissa olevalle funktiolle siten, että kone suoriutui laskennasta. Erityisen merkittävä löytö oli, että Turing osoitti, että koneen nauhalle voitiin koodata mille tahansa tiettyyn ongelmaan spesifioidulle logiikkayksikölle tarkoitetut ohjeet. Tällöin voitiin käyttää yleistä logiikkayksikköä kaikkien laskettavissa olevien ongelmien ratkaisemiseksi. Turing nimesi tällaisen, kaikki laskennalliset ongelmat ratkaisevan, koneen univeraaliksi koneeksi:

”It can be shown that a single special machine of that type can be made to do the work of all. It could in fact be made to work as a model of any other machine. The special machine may be called the universal machine (Turing 1946, 112)”

Vaikka nykypäivän tietokoneissa ei ole Turingin koneen loputonta nauhaa tiedon tallentamiseen, on niiden muistikapasiteetti siinä määrin laaja, että niitä voidaan käytännössä pitää universaaleina. (Nilsson 2010, 58)

Jossain määrin erillään Turingin työstä insinöörit miettivät miten voitaisiin rakentaa laskentaan suunniteltuja laitteita, joiden toiminta pohjautuisi ohjelmiin ja niiden algoritmeja seuraaviin logiikkapiireihin. Vuonna 1937 Claude Shannon osoitti, että boolean algebran ja binääri aritmetiikan avulla voitiin yksinkertaistaa puhelimen piirikytkennän toimintaa. Hän todisti myös, että tuolloin releillä tai tyhjiöputkilla toteutettuja piirikytkentöjä voitiin käyttää myös boolean operaatioiden implementointiin. Sittemmin piirikytkentöjen merkitys tietokoneiden suunnittelussa on ollut merkittävä. (Nilsson 2010, 58-59)

Laskennan nopeutuessa fyysisten komponenttien kehittymisen myötä yhä useampi tehtävä, josta on aiemmin vastannut tehtävään yksilöity kone tai ihminen, voidaan korvata ohjelmointityöllä, jossa universaalille koneelle laaditaan ohjeet kyseisen tehtävän suorittamiseen. Turingin uskoi, että jos kone oli käytännössä universaali, tulisi sen silloin olla kykeneväinen kaikkeen (Turing 1950). Hänen mukaansa oli myös mahdollista, että koneet suorittaisivat toimintoja, joita ei koskaan olisi varsinaisesti ohjelmoitu (Turing 1950). Ensimmäisenä modernina ihmismäisen ajattelun koneellistamista käsittelevänä artikkelina voidaan pitää Turingin 1950-luvulla ilmestynyttä *Computing Machinery and Intelligence* artikkelia (Nilssoin 2010, 61)

Artikkelissaan Turing esittelee koejärjestelyn, jolla voidaan testata koneen päättelykykyä. Turing piti kysymystä voiko kone ajatella liian haastavana, joten hän ehdotti ongelman ratkaisemiseksi sittemmin Turingin testiksi nimettyä matkimistestiä. Turingin testissä on kolme osapuolta: kuulustelija sekä kuulustelun kohteet A ja B, joista toinen on ihminen ja toinen kone. Kuulustelijan on kysymyksiä kysyen selvitettävä kumpi kohteista A vai B on ihminen. Mikäli kone onnistuu hämäämään kuulustelijaa, katsotaan sen ainakin vaikuttavan ajattelevalta ja näin läpäisevän testin. Kuulustelija ei luonnollisesti näe kohteita vaan kommunikoi heidän kanssaan tekstirivikäyttöliittymän välityksellä. (Turing 1950) Vaikka Turingin testiä on kritisoitu puutteelliseksi sekä tasapuolisuutensa, että yksinkertaistamisensa vuoksi, on siitä silti tullut eräs tekoälytutkimuksen klassikoista (Nilssoin 2010, 62).

Liite 2: Oppimistyylit

Murphy (2014, 2) mainitsee oppimistyylien olevan perinteisesti jaettavissa kahteen ryhmään: ohjattuun ja ohjaamattomaan oppimiseen. Ohjatussa oppimisessa koneen tehtävä on oppia vaste y_i annetulle syötteelle x_i käyttäen valmiiksi määriteltyjä syöte-vaste-pareja:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

Missä D on opetusaineisto ja N aineistossa olevien esimerkkitapausten lukumäärä (Murphy 2014, 2). Yksinkertaisimmillaan syöte x_i voi olla moniulotteinen vektori joka kuvaa esimerkiksi aineistossa olevan henkilön pituutta ja painoa. Usein syötteet ovat kuitenkin monimutkaisempia objekteja, kuten kuvia, lauseita, aikasarjoja yms. (Murphy 2014, 2). Haykin (1994) luonnehtii ohjattua oppimista kolmen elementin avulla: ympäristö, opettaja ja opetettava järjestelmä. Ympäristö on informaatiojoukko, josta järjestelmälle opetettava tieto hankitaan ja jalostetaan sopivaan muotoon. Ympäristöstä saatua tietoa kuvataan syöte-vaste-pareina ja käsitellään opettajaelementin välityksellä. Opettaja on määrittelemätön toimija, joka osaa yhdistää ympäristön syötteet virheettömästi oikeisiin vasteisiin ja näin muodostaa pareja, joita voidaan käyttää järjestelmän opetusaineistona. (Haykin 1994) Erilaisia opetettavia järjestelmiä on lukuisia. Suurimmalle osalle on kuitenkin tyypillistä, että vasteen y_i muoto on joko kategorinen (kuten koira tai kissa) tai skalaarinen (kuten esimerkiksi henkilön tulotaso). Vasteen ollessa kategorinen tunnetaan ongelma luokitteluongelmana tai hahmontunnistuksena. Skalaarisen vasteen tapauksessa käytetään regression käsitettä. (Murphy 2014, 3)

Ohjaamaton oppiminen ei tunne Haykinin (1994) mainitsemaa opettajaelementtiä. Ohjaamattomassa oppimisessa järjestelmä sisäistää itsenäisesti ympäristöstä saatavaa informaatiota. Järjestelmälle ei siis tarjota syöte-vaste-pareja opetusaineistoksi vaan aineisto koostuu pelkistä syötteistä:

$$D = \{x_i\}_{i=1}^{N}$$

Missä D on aineisto ja N aineistossa olevien syötteiden lukumäärä. Järjestelmän tavoitteena on löytää datasta mielenkiintoisia säännönmukaisuuksia. (Murphy 2014, 3)

Haykin (1994) jakaa ohjaamattoman oppimisen kahteen suuntaukseen: vahvistusoppimiseen ja itseohjautuvaan oppimiseen. Vahvistusoppiminen määritellään

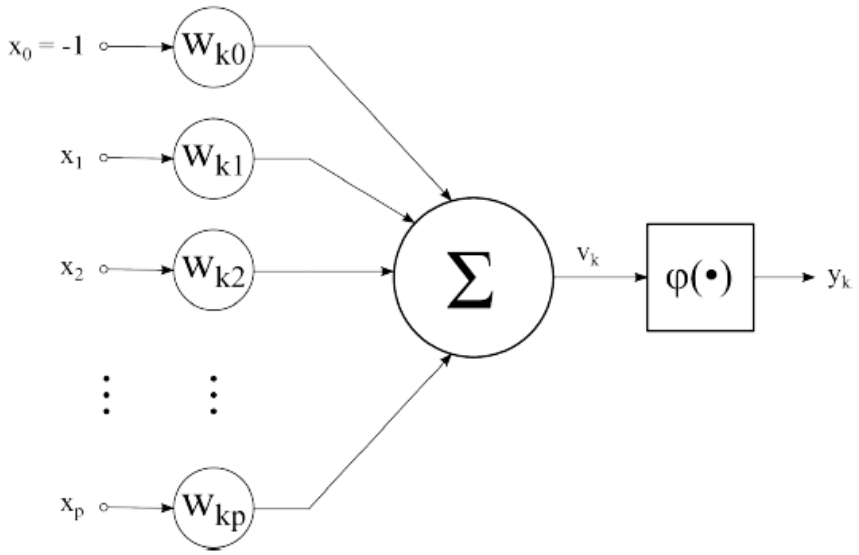
yrityksen ja erehdyksen kautta tapahtuvaksi toiminnoksi. Oppiva järjestelmä oppii suorittamaan toimenpiteitä hyödyntämällä ympäristöstään saamaansa palautetta. (Jang ym. 1997) Kuten ohjatussakin oppimisessa ympäristö mielletään tieto-varannoksi. Opettajan avulla ei kuitenkaan tarjota järjestelmälle syöte-vaste-pareja, vaan järjestelmä saa palautteen suoraan ympäristöltä. Saatu palaute koostuu ärsykkeiden sekvensseistä, joiden avulla järjestelmä muodostaa heuristisia signaaleita, joita kyetään hyödyntämään oppimisprosessissa. Itseohjautuva oppiminen muistuttaa vahvistettua oppimista, mutta heuristisen signaalien muodostumista ei tapahdu ja oppimistyyli on ominainen vain neuroverkolle. Itseohjautuvassa oppimisessa järjestelmän vapaat parametrit muunnetaan ympäristön tuottamien syötteiden mukaisiksi. Järjestelmä optimoituu ympäristön tiedolle ja pystyy näin poimimaan piirteitä syötteistä. (Haykin 1994) Toisin kuin ohjatussa oppimisessa ohjaamattoman oppimisen ongelman asettelu on hyvin väljä, sillä etukäteen ei tiedetä mitä järjestelmälle syötettävästä datasta tulisi tunnistaa (Murphy 2014, 3) Louridas ja Ebert (2016) havainnollistavat oppimistyylejä vertaamalla niitä opiskelijalle annettuihin erilaisiin tehtäviin. Ohjatussa oppimisessa opiskelijalle annetaan sekä ongelma että sen ratkaisu ja pyydetään tämän jälkeen ratkaisemaan muita saman tyylisiä ongelmia. Ohjaamattomassa oppimisessa opiskelija saa aineiston ja häntä pyydetään löytämään sääntöjä ja selityksiä aineiston muodostumiselle. (Louridas ja Ebert 2016)

Liite 3: Neuroverkot

Neuroverkon pienin omaksi entiteetikseen erotettava kokonaisuus on neuroni. Keinotekoisen neuronin tapa vastaanottaa syötteitä ja levittää niitä eteenpäin jakaa tiettyjä ominaisuuksia elollisten olentojen aivoissa olevien neuronien kanssa (Fausett 1994). Haykin (1994, 8) jakaa neuronin kolmeen peruselementtiin:

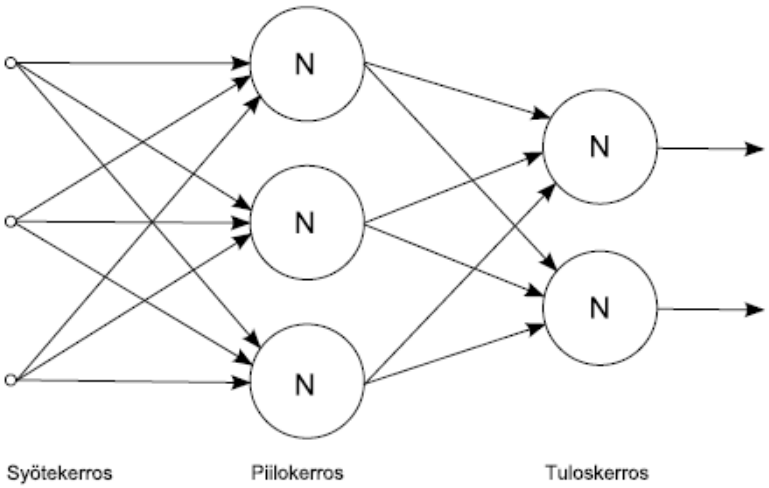
1) *Synapsien joukko* on numeerisesta datasta muodostuvia syötteitä, jotka saapuvat neuronille joko ympäröivästä maailmasta, tai edeltävän kerroksen neuroneilta. Jokaiseen synapsiin liittyy oma painoarvokerroin, jolla saapuva numeerinen data kerrotaan.
2) *Summaaja* on elementti, joka laskee neuroniin saapuvat painoarvokertoimilla kerrotut syötteet yhteen.
3) *Aktivaatiofunktio* on funktio, jonka tehtävänä on määrittää syötetiedon ja tuloksen keskinäinen riippuvuus. Haykin (1994, 8) määrittelee aktivaatiofunktion rajoittimeksi, jonka tehtävänä on rajata neuronin lähettämä vaste tietylle arvoalueelle. Yleensä käytössä olevat aktivaatiofunktiot saavat arvoja joku suljetulla välillä [0,1] tai [-1,1]. Neuroverkolle annettu tehtävä määrittää aktivaatiofunktion valinnan. (Haykin 1994,8) Eri aktivaatiofunktioiden tarkempi käsittely ei tässä tutkimuksessa ole tutkimuskysymyksen kannalta tarkoituksenmukaista.

Kuviossa 11 on havainnollistettu neuronin rakennetta. $x_0, x_1 \dots x_p$ kuvaavat neuronin saamia numeerisia syötesignaaleja (*synaptinen joukko*), joista jokaisella on oma painoarvonsa $w_{k0}, w_{k1} \dots w_{kp}$. Merkillä Σ kuvataan *summaajaa* ja $\varphi(\cdot)$ on *aktivaatiofunktio*, joka rajoittaa neuronin eteenpäin välittävän tuloksen y_k halutulle välille.



Kuvio 11: Neuroni (Haykin 1994, 8 mukaillen)

Pelkän yksittäisen neuronin antama vaste y_k riittää harvoin ratkaisemaan neuroverkon työstettäväksi annettua käytännön ongelmaa. Neuroverkot toteutetaankin usein monen neuronin yhdistelmänä. Erilaisia neuroverkkorakenteita tunnetaan useita. Syöte- ja tuloskerroksen olemassaolo on kuitenkin yhteistä kaikille verkkorakenteille. Useimmiten verkkorakenne koostuu syötekerroksesta, yhdestä tai useammasta piilokerroksesta sekä tuloskerroksesta. Syötekerros vastaanottaa syötteet ja välittää ne eteenpäin ensimmäisen piilokerroksen neuroneille. Piilokerroksen neuronit välittävät tuloksensa aina eteenpäin seuraavalle piilokerrokselle, kunnes saavutetaan tuloskerros. Edellä kuvattua vasteiden välitystä kutsutaan eteenpäinsyöttämiseksi. Tuloskerroksen vaste vastaa muodoltaan verkolle annettua ongelmaa. (Haykin 1994, 18-22) Esimerkiksi luokitteluongelmassa, jossa verkon tulisi erotella kissojen kuvat koirien kuvista, tuloskerros voitaisiin määritellä koostumaan yhdestä neuronista, joka antaisi todennäköisyyden sille, että syötteeksi annettu kuva esittää koiraa, mikäli neuronin antama todennäköisyys olisi alhainen, voitaisiin päätellä, että kuvassa on kissa. Yhdestä piilokerroksesta koostuva eteenpäinsyöttävä verkko on esitetty kuviossa 12



Kuvio 12: Yhden piilokerroksen eteenpäinsyöttävä neuroverkko

Haykin (1994, 18-22) jakaa yleisimmät neuroverkkorakenteet neljään eri kategoriaan: *yksi- ja useampikerroksisiin eteenpäinsyöttäviin verkkoihin, toistuviin verkkoihin sekä hilaverkkoihin.* Eri verkkorakenteet eroavat toisistaan kerroksien lukumäärän, syöte-ja tuloskerroksen järjestelyn sekä tuloskerroksen vasteen käsittelyn suhteen. Käsiteltävä ongelma määrittää mitä verkkorakennetta käytetään.

Liite 4: K:n lähimmän naapurin menetelmä

Erilaisia lähimmän naapurin menetelmiä käytetään tyypillisesti erityisesti luokittelutehtävissä ja ne soveltuvat lähes kaikenlaisen datan luokitteluun. Lähimmän naapurin menetelmässä datan luokittelu tehdään valittuun etäisyysmittaan pohjautuen ja algoritmi on jaettavissa neljään eri vaiheeseen.

1) Valitse vertailtavien naapurien lukumäärä K
2) Ota tarkasteluun luokiteltavan datapisteen K lähintä datapistettä eli naapuria
3) Lake kuinka monta naapuria kuuluu mihinkäkin kategoriaan.
4) Määritä luokiteltava datapiste kuuluvaksi samaan kategoriaan kuin suurin osa sen naapureista.

Etäisyysmitan valinta on lähimmän naapurin menetelmän kriittisin vaihe. Käytännössä etäisyysmitta voi olla mikä tahansa metodi, jolla kahta datayksikköä voidaan verrata toisiinsa ja selvittää kuinka lähellä toisiaan ne ovat. Etäisyysmitan tulee kuitenkin täyttää seuraavat neljä ominaisuutta: 1) etäisyyden tulee olla aina positiivinen 2) etäisyys on nolla, vain mikäli vertailussa olevat havainnot ovat identtiset 3) etäisyys on sama riippumatta havaintojen järjestyksestä 4) mikäli kahden havainnon sijasta käytetään kolmea havaintoa, etäisyyden tulee olla suurempi kuin kahta havaintoa käytettäessä.

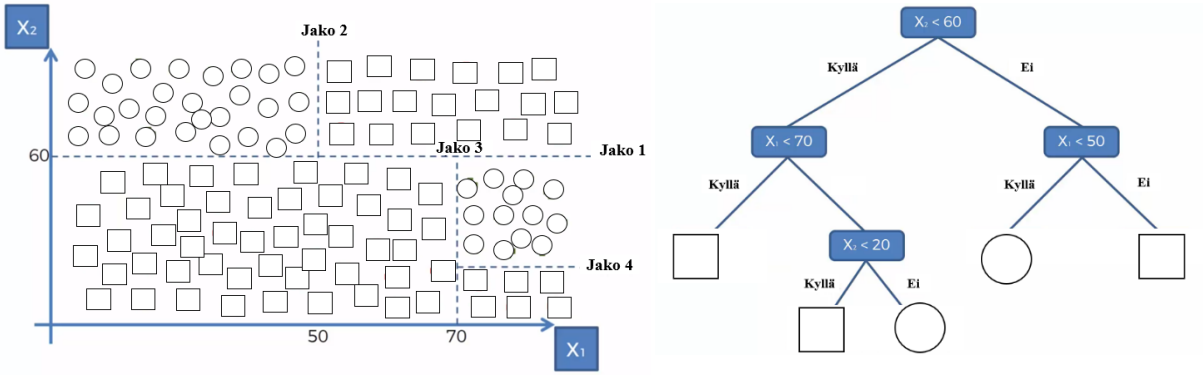
Datan laatu määrää menetelmässä käytettävän etäisyysmitan. Esimerkiksi jatkuville muuttujille voidaan hyödyntää euklidista etäisyyttä. Tällöin etäisyys d voidaan laskea seuraavasti:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

jossa n on koordinaattien eli eri ulottuvuuksien lukumäärä, x_i on pisteen x arvo koordinaatissa i ja y_i pisteen y arvo koordinaatissa i.

Liite 5: Päätöspuut

Yleisellä tasolla päätöspuut ovat työkalu päätöksenteon kuvaamiseen puurakenteen avulla. Päätöspuun kukin oksa kuvaa yhtä havaintoa ja kukin lehtisolmu lopputulosta. Koneoppimisessa päätöspuu toimii ennakoivana, opetusdatan pohjalta luotavana, mallina. Tällöin oksat kuvaavat datapisteestä tehtyjä havaintoja ja lehtisolmut johtopäätöksiä datapisteen siitä muuttujasta, jota päätöspuun avulla halutaan ennakoida. (Maimon ja Rokach, 2008) Päätöspuita voidaan soveltaa sekä regressio, että luokitteluongelmiin. Kuviossa 13 on havainnollistettu luokitteluongelmaan opetettavan puun toimintaa



Kuvio 13: Päätöspuu (Eremenkoa 2018 mukaillen)

Päätöspuuta luotaessa datasta valitaan joillain kriteereillä ominaisuus, jonka perusteella data jaetaan osiin. Tätä toistetaan kullekin jaossa syntyneelle datan osalle, kunnes kaikki opetusdatan alkiot on luokiteltu tai jaottelun tuloksena syntyneissä osajoukoissa on liian vähän alkioita ennalta määritettyyn minimiarvoon nähden. (Maimon ja Rokach, 2008) Kuviossa 13 jakokriteerinä toimii silmämääräisellä tarkastelulla erotettavissa oleva geometrinen muoto ja jako toteutetaan kahden ennakoivan muuttujan x_1 ja x_2 suhteen. Kuvion 13 oikealla puolella on jaon tuloksena syntynyt päätöspuu.